NATIONAL CENTER FOR EDUCATION STATISTICS

Working Paper Series

The Working Paper Series was initiated to promote the sharing of the valuable work experience and knowledge reflected in these preliminary reports. These reports are viewed as works in progress, and have not undergone a rigorous review for consistency with NCES Statistical Standards prior to inclusion in the Working Paper Series.

U. S. Department of Education
Institute of Education Sciences

This page intentionally left blank.

NATIONAL CENTER FOR EDUCATION STATISTICS

Working Paper Series

IMPUTATION METHODOLOGY FOR THE NATIONAL POSTSECONDARY STUDENT AID STUDY: 2004

Working Paper No. 2003-20 August 2003

Contact: James Griffith
Project Officer
James.Griffith@ed.gov

U. S. Department of Education
Institute of Education Sciences

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

> National Center for Education Statistics
> Institute of Education Sciences
> U.S. Department of Education
> 1990 K Street NW
> Washington, DC 20006–5651

August 2003

The NCES World Wide Web Home Page address is *http://nces.ed.gov*
The NCES World Wide Web Electronic Catalog is: *http://nces.ed.gov/pubsearch*

Suggested Citation

U.S. Department of Education, National Center for Education Statistics. *Imputation Methodology for the National Postsecondary Student Aid Study: 2004*. NCES 2003–20, by Kimberly Ault, Stephen Black, Jim Chromy, Mansour Fahimi, Peter Siegel, Lily Trofimovich, Roy Whitmore, and Lutz Berkner. Project Officer: James Griffith. Washington, DC: 2003

For ordering information on this report, write:

> U.S. Department of Education
> ED Pubs
> P.O. Box 1398
> Jessup, MD 20794–1398

Or call toll free 1–877–4ED–Pubs

Content Contact:
James Griffith
(202) 502–7387
James.Griffith@ed.gov

Foreword

In addition to official NCES publications, NCES staff and individuals commissioned by NCES produce preliminary research reports that include analyses of survey results, and presentations of technical, methodological, and statistical evaluation issues.

The ***Working Paper Series*** was initiated to promote the sharing of the valuable work experience and knowledge reflected in these preliminary reports. These reports are viewed as works in progress, and have not undergone a rigorous review for consistency with NCES Statistical Standards prior to inclusion in the Working Paper Series.

Copies of Working Papers can be downloaded as pdf files from the NCES Electronic Catalog (http://nces.ed.gov/pubsearch/), or contact Sheilah Jupiter at (202) 502–7363, e-mail: sheilah.jupiter@ed.gov, or mail: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 1990 K Street NW, Room 9048, Washington, DC 20006.

Marilyn M. Seastrom
Chief Mathematical Statistician
Statistical Standards Program

Ralph Lee
Mathematical Statistician
Statistical Standards Program

This page intentionally left blank.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This report focuses on issues relating to data editing and imputation methodology for the National Postsecondary Student Aid Study (NPSAS). This chapter gives a brief description of the NPSAS, outlines the study objectives, and discusses the proposed imputation methodology. Chapters 2 through 5 discuss each of the four study objectives that will improve the speed, efficiency, and reliability of the data editing and imputation process for the next NPSAS in 2004. Chapter 6 summarizes the results of the four different study objectives and states recommendations for implementing the proposed methods.

Background

The NPSAS is a comprehensive study of financial aid among postsecondary education students in the United States and Puerto Rico, and provides current information on how families pay for postsecondary education. The study is based on a nationally representative sample of students in postsecondary education institutions, which includes undergraduate, graduate, and first-professional students. Represented are students attending all types and levels of institutions, including public, private not-for-profit, and private for-profit institutions, and institutions ranging from less-than-2-year institutions to 4-year colleges and universities.

The data analysis files for NPSAS:2000 contain student-level data for about 50,000 undergraduates and about 12,000 graduate and first-professional students who were enrolled in institutions of postsecondary education during the 1999–2000 academic year. These sample members include about 600 variables that were either based directly on or derived as composites from a variety of different sources. The most important sources are the following:

- CADE (computer-assisted data entry) – Contains raw data collected from institutional records. The data include some basic student characteristics (birth date, gender, marital status, race/ethnicity, and citizenship), attendance and educational program information, dependency status, financial aid, tuition, and student budgets. CADE institutional data are available for nearly all CATI respondents in NPSAS:2000.
- CATI (computer-assisted telephone interview) – Contains raw data collected from students who responded to the telephone interview. The data include many of the same variables describing student characteristics, attendance, and educational programs as well as additional information that is not usually collected by institutions (parental education, family income, family size, student employment, financial aid received from outside sources, loan debt, disabilities) and questions about student attitudes and personal experiences (voting behavior, community service activities, credit card use). In NPSAS:2000, about 75 percent of the study members completed the full CATI interview and another 5 percent completed an abbreviated interview.

- CPS (Central Processing System) – Contains data from the federal database of financial aid applications filed in 1999–2000. The financial aid applications include detailed information about student and parental demographics (income and assets, marital status, family size, parental education, citizenship) and the calculated expected family contribution and dependency status that are used in financial aid need analysis. About one-half of the study respondents had financial aid application data in this file.

Sample cases that had relatively complete data from any two of these sources were considered study respondents in NPSAS:2000 because most of the variables needed to address the most important objectives of the study (college costs and student aid) could be derived from any two of them. Most of the data can be analyzed using the study weight based on all respondents. However, for types of information available only through the student interview (employment, attitudes, experiences), a separate CATI weight was calculated that adjusted for the 25 percent of cases that did not include a full CATI interview.

Study Objectives

The new "statistical standards" adopted by the National Center for Education Statistics (NCES) in September 2002 include Standard 4-1-2, which states "key variables in data sets used for cross-sectional estimates <u>must</u> be imputed." The NCES Postsecondary Studies Division has interpreted "key variables" to be those that are used in a "releasing publication," which for NPSAS includes the E.D. Tab, the methodology report, a descriptive summary, and two finance reports (for undergraduate students and for graduate/first-professional students). Therefore, the number of variables that must be imputed will be much more than have ever been imputed in past NPSAS studies. In addition, the imputation work must be completed quickly because of the contractual requirement to produce deliverable files and reports on a much tighter schedule than ever before achieved (e.g., adjudication-ready methodology report by November 30, 2004, and Data Analysis System [DAS] by December 15, 2004). Therefore, this research project was undertaken to develop imputation procedures that can be implemented quickly and efficiently for NPSAS:2004.

The overall objective of this research effort is to develop methods for NPSAS:2004 that will improve the speed, efficiency, and reliability of the data editing and imputation processes. Specific study objectives of this research are as follows:

1. Develop a procedure for resolving discrepancies among the different data sources. For several demographic variables, the effect of reducing the logical imputation effort by defining certain conflicting data as missing and then imputing those data using stochastic imputation procedures has been tested.
2. Develop a quick and efficient method for imputing variables with low levels of missing values. Low levels of missing values are common for many CATI interview items.
3. Develop a more efficient method for imputing values when information is only available for some classes of respondents. One of the most important of these is the price of attendance (student budget), which is only available for those students who applied for financial aid (about one-half of the sample).

4. Estimate the relative magnitude of the imputation component of variance for a small number of variables.

In the next NPSAS (2004), student records will only be included as study respondents if they have available both relatively complete institutional CADE and student CATI interview data (no abbreviated interviews will be conducted). Therefore, the imputation methodology developed for this study used the NPSAS:2000 student records that had full CATI interview and CADE data and was limited to the undergraduates (about 60 percent of the total cases). The NPSAS:2000 data were used for this research because they were the most recent data available.

Imputation Methodology

Weighted sequential hot deck imputation was the procedure used to support all study objectives (Cox 1980 and Iannacchione 1982). Sequential hot deck imputation is a common procedure used for item nonresponse. This method uses the respondent survey data as donors to provide surrogate values for records with missing data. The basic principle of sequential hot deck imputation involves defining imputation classes, which generally consist of a cross-classification of covariates, and then replacing missing values sequentially from a single pass through the survey data within the imputation classes. Weighted sequential hot deck imputation uses the sampling weights of the item respondents and nonrespondents. This procedure takes into account the unequal probabilities of selection in the original sample by using the sampling weights to specify the expected number of times a particular respondent's answer will be used to replace a missing item. These expected selection frequencies are specified so that, over repeated applications of the algorithm, the expected value of the weighted distribution of the imputed values will equal in expectation within imputation class the weighted distribution of the reported answers.

Weighted sequential hot deck imputation has as an advantage that it controls the number of times a respondent record can be used for imputation and gives each respondent record a chance to be selected for use as a hot deck donor. To implement the weighted hot deck procedure, imputation classes and sorting variables that were relevant for each item being imputed were defined. If more than one sorting variable was chosen, a serpentine sort was performed where the direction of the sort (ascending or descending) changed each time the value of the previous sorting variable changed. The serpentine sort minimized the change in the student characteristics every time one of the sorting variables changed its value.

Imputation classes were developed by using a Chi-squared automatic interaction detection (CHAID) analysis where only the respondent data were modeled. The CHAID segmentation process first divides the data into groups based on categories of the most significant predictor of the item being imputed, and then splits each of these groups into smaller subgroups based on other predictor variables. The CHAID process may also merge categories of a variable that were found not to be significantly different. This splitting and merging process continued until no more statistically significant predictors were found (or until some other stopping rule was met). Imputation classes were defined from the final CHAID segments.

This page intentionally left blank.

Chapter 2

Resolving Discrepancies Between Data Sources

Statement of Objective

The first study objective was to develop a procedure for resolving inconsistencies for the same or similar data items among the multiple data sources in NPSAS (e.g., institutional [CADE], CATI student interview, and federal database [CPS] records). In the NPSAS:2000, when the same variables were available from more than one source but had different values, logical imputation (such as assigning priorities for the choice of sources) was used to resolve all inconsistencies. This process is typically labor intensive because it involves a close examination of the data and development of a separate procedure for many different data patterns or scenarios. Often after logical checks and assignments are made, many cases are still unresolved. When data editing (logical imputation) fails to resolve inconsistencies, an attractive alternative is to declare such data items to be missing and then fill in the missing information via stochastic imputation. This is the proposed procedure tested and discussed in this chapter.

The following four test variables were selected to compare the difference between the results using logical imputation (the main procedure actually used in the NPSAS:2000) and reliance on more extensive stochastic imputation:

- Student citizenship (CITIZEN2);
- Student marital status (SMARITAL);
- Whether the student had any dependents (ANYDEP); and
- Number of dependents (NDEPEND).

Methodology

The programming code for the logical imputation of these variables in NPSAS:2000 was reviewed and rewritten to remove all logical imputation used to resolve differences in conflicting data among sources. Whenever there were data source discrepancies, the value was set to missing. In addition, values that had previously been imputed stochastically (because no source was available) were set to missing. The weighted percentage of values reset to missing ranged from about 3 percent for citizenship to about 10 percent for marital status. Table 1 shows the distributions for each of the four test variables before any imputation, including the percentage with completely missing data (that had been imputed stochastically in NPSAS:2000) and the percentage set to missing retroactively because of conflicting data sources.

In NPSAS:2004, inconsistencies that can easily be resolved (e.g., because two of three sources agree or because one source is considered to be most reliable) will be resolved through logical imputations. The remaining inconsistencies that cannot be easily resolved are candidates to be set to missing and imputed using the procedures being reported herein. For this test, *all* discrepant data were set to missing.

Table 1. Distribution of test variables before imputation

Variable	Sample size[1]	Weighted percent	Weighted known percent
Student citizenship (CITIZEN2) total	**32,780**	**100.0**	**100.0**
U.S. citizen or U.S. national	29,830	89.5	95.4
Resident alien	800	3.1	3.3
Foreign/international student	350	1.3	1.4
Missing	970	3.4	
Conflicting information	840	2.7	
Student marital status (SMARITAL) total	**32,780**	**100.0**	**100.0**
Single	23,880	71.7	78.2
Married	5,530	19.3	21.0
Separated	210	0.8	0.9
Missing	20	0.1	
Conflicting information	3,140	8.1	
Any dependents (ANYDEP) total	**32,780**	**100.0**	**100.0**
Had no dependents	24,140	71.3	74.6
Had dependents	7,020	24.3	25.4
Missing	70	0.3	
Conflicting information	1,550	4.1	
Number of dependents (NDEPEND) total	**32,780**	**100.0**	**100.0**
0	24,130	71.3	74.8
1	2,580	8.7	9.1
2	2,470	8.8	9.2
3	1,210	4.3	4.5
4 or more	690	2.3	2.4
Missing	300	0.9	
Conflicting information	1,400	3.8	

[1]Includes sampled undergraduate students who had a full computer-assisted telephone interview (CATI).
NOTE: Detail may not add to total due to rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), Restricted use data file.

Prior to imputation of the test variables, several variables with low levels of missing values were imputed to help define imputation classes for the test variables. These initial imputations were done using weighted sequential hot deck imputation within imputation classes defined by age and gender and using the following sorting variables: region, institution control (public, private not-for-profit, private for-profit), and institution level (4-year, 2-year, less-than-2-year). These variables and the percentage of missing values that were imputed are shown in table 2 below.

Table 2. Variables imputed prior to test variables

Variable	Percent missing
Student U.S. born (NBUSBORN)	0.2
Student's primary language (NBLANG)	0.3
Student's parents foreign born (PARBORN)	1.0
Hours student worked (WORKED)	1.3
Student had job while enrolled (ENRJOB)	2.5
Primary student role (SEROLE)	3.1

NOTE: The percent missing was calculated using cases for sampled undergraduate students who had a full computer-assisted telephone interview (CATI).
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), Restricted use data file.

As stated in chapter 1, the imputation method used was a weighted sequential hot deck procedure, and a CHAID analysis was used to define imputation classes. The test variables were imputed in the order as shown in table 1. The order of imputation addressed problems of multivariate association by using a series of univariate models fitted sequentially such that variables modeled earlier in the hierarchy had a chance to be included in the covariate set for subsequent models. Appendix A shows the results from the CHAID analysis for each of the test variables.

Results

Often the most difficult part of a comparison of different methods is developing sound evaluation criteria. To determine the similarity of the after imputation distributions, absolute differences and relative percent differences between the proposed and prior imputation procedures were examined. The absolute differences were calculated by subtracting the prior method weighted percent from the proposed method weighted percent (displayed as *proposed – prior* in equation below). The relative percent differences were calculated by dividing the difference of the two values by their average, or:

$$Relative\ Percent\ Difference = \frac{proposed - prior}{(proposed + prior)/2} * 100\%$$

Relative percent differences less than or equal to 10 percent were considered acceptable results and relative percent differences greater than 10 percent were examined more closely to determine reasons for the differences. Distributions were considered to be similar when absolute differences were less than 2 percent. Table 3 shows the after imputation distribution of the test variables using the NPSAS:2000 imputation method and the proposed imputation method. The relative percent differences between the levels of the variable distributions ranged from 0 percent to about 40 percent.

The distributions of test variables were tested in SUDAAN to determine whether the proposed imputation method produced significantly different results from the prior NPSAS:2000 method. This test was performed to provide a distributional overview of each test variable and to determine if there were significant differences among imputation methods. T-tests were used to test each level of the test variables for differences at the 0.05/(c-1) significance level, where "c" is the number of categories within the variable. The t-tests were performed to give a better overall picture of what was occurring between imputation methods for each test variable. The t-tests provided a clearer picture of what was occurring because they were performed at each level of the test variables, which provided further evidence as to the significance of the overall distribution of the test variables. The t-tests were computed using SUDAAN. For the first three variables (student citizenship, student marital status, and any dependents), the variable distributions showed significant differences overall and significant differences at all levels. The distribution of the number of dependents variable showed a significant difference overall and for the first three levels (0, 1, and 2). However, it did not show significant differences for the remaining levels (3 and 4 or more).

For the variable student citizenship, the largest relative percent differences were for the noncitizen categories. In the logical imputation procedure for the NPSAS:2000, if student citizenship data sources had conflicting information, then the data source values that maximized noncitizenship were assigned. The weighed sequential hot deck procedure distributed the missing cases according to the respondent data distribution. Therefore, the significant differences detected can be attributed to the effects of the logical imputation procedure.

For the variable student marital status, the largest percent difference was for the "separated" category. Appendix B shows the relative percent differences between the prior and proposed imputation procedures by imputation classes for student marital status. The largest relative percent differences were found in the following imputation classes for the "separated" category:

- In CPS, male students between 24 and 29 years old who did not work while enrolled,
- In CPS, female students between 24 and 29 years old who did not work while enrolled,
- In CPS, students between 24 and 29 years old who worked part time while enrolled, and
- In CPS, students between 24 and 29 years old who worked full time while enrolled.

The NPSAS:2000 imputation method for student marital status used 10 imputation classes formed from the following variables: CPS record indicator, fall enrollment status, student type, and age. The proposed method used the same variables and one additional variable – work status while enrolled. Most of the differences were attributed to this additional imputation class variable.

Table 4 shows the distribution of the other data files from which the student citizenship (CITIZEN2) and marital status (SMARITAL) variables were derived. For student citizenship, the CATI and CADE variable distributions were similar (less than 2 absolute difference in the subcategory weighted percentages) to the postimputation distribution for both imputation methods (see table 3.) For student marital status, the CATI variable distribution was similar to the postimputation distribution for both imputation methods.

Table 3. Distribution of test variables after imputation, by method

Variable	Prior imputation method		Proposed imputation method		Absolute difference	Relative percent difference
	Sample size[1]	Weighted percent	Sample size	Weighted percent		
Student citizenship (CITIZEN2) total	**32,780**	**100.0**	**32,780**	**100.0**		
U.S. citizen or U.S. national	30,960	93.4	31,240	94.3	0.9	1.0
Resident alien	1,350	4.7	1,140	4.1	0.6	–13.6
Foreign/international student	480	1.9	410	1.6	0.3	–17.1
Student marital status (SMARITAL) total	**32,780**	**100.0**	**32,780**	**100.0**		
Single	25,640	76.0	26,330	77.7	1.7	2.2
Married	6,660	22.6	6,220	21.4	1.2	–5.5
Separated	490	1.4	240	0.9	0.5	–43.5
Any dependents (ANYDEP) total	**32,780**	**100.0**	**32,780**	**100.0**		
Had no dependents	24,650	72.5	25,350	74.6	2.1	2.9
Had dependents	8,130	27.5	7,440	25.4	2.1	–7.9
Number of dependents (NDEPEND) total	**32,780**	**100.0**	**32,780**	**100.0**		
0	24,650	72.5	25,350	74.6	2.1	2.9
1	3,320	10.8	2,760	9.2	1.6	–16.0
2	2,760	9.5	2,640	9.2	0.3	–3.2
3	1,310	4.6	1,310	4.6	#	#
4 or more	740	2.5	730	2.4	0.1	–4.1

Rounds to zero.
[1]Includes sampled undergraduate students who had a full computer-assisted telephone interview (CATI).
NOTE: All subcategories for the variables student citizenship, student marital status, and any dependents showed significant differences. For the number of dependents, the first three subcategories showed significant differences. Detail may not add to total due to rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), Restricted use data file.

The large percent difference for the "separated" marital status category is the most noticeable difference between the two methods. However, the percentage of cases falling into the "separated" category (0.9 percent) for the proposed imputation method is essentially the same as the percentage of cases falling into this same category for both the CATI and CADE variables (also 0.9 percent).

Conclusions

One of the reasons for discrepancies among the data sources for these variables is that valid responses may change over time, and the sources represent different points in time. In general, the financial aid application (CPS source) is filed in the spring before the beginning of the NPSAS sample year; the institutional data (CADE) represent registrar records that are usually filed in the fall of the NPSAS year; and the student interview (CATI) is conducted the following spring or summer. Therefore, there may be a difference of up to 18 months in the sources. A student's citizenship, marital status, and number of dependents can change during this period.

This is especially evident in the “separated” category for marital status. Separation is a transitory condition between marriage and divorce, when the student is single again, so it is not surprising that the distribution changes when the assumptions used in the logical imputation are different from those in the stochastic imputation.

The reasons for the significant differences in the variable distributions between methods can be attributed to the prior logical imputation procedure used and the differences in imputation classes. There are no criteria for judging which of the two methods is “better,” because it is unknown whether the discrepancies represent valid changes in status that take place over time or whether they represent reporting errors. Therefore, the stochastic imputation method, which is more efficient (because it requires minimal data analysis) and can be implemented quickly, provides results that are just as acceptable as the logical imputation method. This procedure meets the criteria for improving the speed, efficiency, and reliability of the data editing and imputation process.

Table 4. CADE, CATI, and CPS data source variables

	Data source					
	CATI[1]		CPS[2]		CADE[3]	
Variable	**Sample size[4]**	**Weighted percent**	**Sample size**	**Weighted percent**	**Sample size**	**Weighted percent**
Student citizenship total	**32,780**	**100.0**	**32,780**	**100.0**	**32,780**	**100.0**
U.S. citizen or U.S. national	31,380	94.8	18,030	93.7	28,590	94.0
Resident alien	990	3.7	980	6.2	980	4.1
Foreign/international student	390	1.5	10	0.1	470	1.9
Missing/not applicable	30		13,760		2,740	
Student marital status total	**32,780**	**100.0**	**32,780**	**100.0**	**32,780**	**100.0**
Single	24,350	76.1	15,650	82.8	17,990	82.0
Married	7,920	23.1	2,960	15.0	3,580	17.1
Separated	440	0.9	410	2.2	190	0.9
Missing/not applicable	80		13,760		11,020	

[1]CATI=computer-assisted telephone interview.
[2]CPS=Central Processing System.
[3]CADE=computer-assisted data entry.
[4]Includes sampled undergraduate students who had a full CATI interview.
NOTE: Detail may not add to total due to rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), Restricted use data file.

Chapter 3

Imputation Procedures for Variables with Low Levels of Missing Values

Statement of Objective

NCES Statistical Standard 4-1-2 requires that all "key variables" be imputed. Therefore, it is anticipated that the number of variables requiring imputation in the NPSAS:2004 study may be much greater than in past studies. However, because study respondents in NPSAS:2004 will be defined as students having relatively complete information from both institutional (CADE) and student interview (CATI) sources, it is anticipated that there will be numerous items with relatively low levels of missing values (i.e., less than 10 percent). The second study objective was to develop a quick and efficient imputation procedure that could be used for any variable that had a low percent of missing values.

Methodology

For this objective, a group of variables in NPSAS:2000 that had low levels of missing values for cases that had CATI data were examined. Table 5 shows the variables selected for this objective. These variables had less than 10 percent missing values and were based on the student CATI interview. The weighted sequential hot deck imputation procedure was chosen because it can be performed quickly on a large set of variables when the data files are not extremely large and when imputation classes are readily available.

Typically, an analysis of the data being imputed is necessary to find predictor variables and imputation classes. To find a set of predictor variables for each imputation variable, a CHAID analysis is performed on all potentially correlated variables on the data file. Because the number of variables in the data files is quite large, the time and effort required to find an efficient set of highly correlated predictor variables for each variable to be imputed outweighs any significant differences in the final distributions when the percentage of missing values is low. If a "predetermined set" of variables is selected as potential predictors for a CHAID analysis, then imputation classes can be defined quickly and will be available for the weighted sequential hot deck imputation procedure. Further efficiency can be gained if a "predetermined set" of imputation classes is considered sufficient.

The hierarchical approach for imputation, where prior imputed variables are used in subsequent models, is an effective method when imputation variables are related. However, the process of selecting imputation classes for each subsequent variable has to be performed and this repeated process of imputing a variable, running a CHAID analysis, can be quite time consuming.

Therefore, the procedure tested in this objective used a predetermined set of predictor variables and did not use the hierarchical approach. For each test variable, a CHAID analysis was performed on the set of predictor variables and then each variable was imputed using weighted sequential hot deck imputation. The predetermined set of variables was chosen by the following criteria: institution- or student-related characteristic variables and predictor variables used in the NPSAS:2000 imputation method. The list of predictor variables shown in table 5 was the only set tested; however, other variables could be added or some could be excluded. Some predictor variables listed in table 5 may be redundant (e.g., OBE region and institution state). However, the CHAID analysis chooses the most appropriate set of variables, regardless of redundancies.

Table 5. Test variables with low levels of missing values and predictor variables

Test variables	Weighted percent missing
Ever vote (NBEVRVT)	0.6
Hours worked per week in NPSAS year (NDHOURS)	2.2
Number of credit cards in own name (NDNUMCRD)	3.9
Mother's education (NBMOMED)	4.9
Amount of commercial loan (NCAMNCOM)	5.5
Father's education (NBDADED)	8.4
Amount of employer aid (NCAMNEMP)	9.3
Predictor variables	
Age as of 12/31/1999 (AGE)	
Gender (GENDER)	
Race category (RACECAT)	
Student dependency status (DEPEND)	
Attendance intensity in fall (ATTEND)	
Student's citizenship (CITIZEN2)	
Hispanic ethnicity (HISPANIC)	
Student's marital status (SMARITAL)	
Student housing (LOCALRES)	
NPSAS sample institution level (LEVEL)	
NPSAS sample institution type (CONTROL)	
Student U.S. born (NBUSBORN)	
Institution state (INSSTATE)	
OBE[1] region code (OBEREG)	
CPS[2] data indicator (INCPS)	
Applied for federal aid (FEDAPP)	
Degree of urbanization (LOCALE)	
Respondent had any dependents (ANYDEP)	

[1]OBE region=Bureau of Economic Analysis Region Code.
[2]CPS=Central Processing System.
NOTE: The percent missing was calculated using cases for sampled undergraduate students who had a full computer-assisted telephone interview (CATI). Study respondents were placed in one of five race categories, one category for each of the five races. Whenever a study respondent's response was some multiple configuration of races, the most "minority race" (the one race within the configuration with the fewest respondents) was assigned. This hierarchy, from most "minority" to least, was American Indian, Pacific Islander, Asian, Black, and White.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), Restricted use data file.

For this objective, it was assumed that any variable listed as a potential predictor variable had no missing values or had already been imputed. A standard procedure for imputation is to impute key demographic variables, such as age, gender, and race, which typically have low levels of missing values, then use these variables to help impute other variables in a sequential order. Additionally, variables may be ordered by the percent missing and imputed sequentially from the lowest to the highest level. The procedure tested in this chapter assumed that key demographic variables with low levels of missing values would be imputed first (similar to the variables examined in chapter 2) and then the "predetermined set" would be chosen. Finally, the large number of variables with low levels of missing values would be imputed.

To address the speed issue, the CHAID analysis and the imputation procedures were integrated. The imputation classes shown in appendix C were developed using the SPSS Answer Tree software, a learning system that creates a collection of classification rules displayed in decision trees. The classification rules were exported as a SAS-like code file, and with simple modification, the code was used in a SAS program. After the imputation classes were generated, SAS was used to perform the weighted sequential hot deck imputation. The procedure of exporting the decision tree results from SPSS and importing them to SAS is almost automated so that the imputation procedure can be performed relatively quickly. Complete automation, however, was not possible with Answer Tree, so other software packages were considered. These included CART, KnowledgeSTUDIO, and SAS user-written macros called CHAID and TREEDISC. Due to the limited time available for this research, alternatives to Answer Tree were not fully explored. If more time is allocated to explore these other options, considerable enhancement of the process of integrating the CHAID and imputation procedures may be possible.

Results

Appendix C shows the imputation classes resulting from the CHAID analysis and the sorting variables used in the imputation procedure. As shown in appendix C, the resulting imputation classes for mother's and father's education are different. The CHAID analysis selected the following covariates for father's education—student dependency status and financial aid application status—and the following covariates for mother's education—student dependency and Hispanic status. Regarding the issue of redundancy of predictor variables, the CHAID results did not choose the OBE region nor the institution state as potential predictors as part of the final imputation classes. However, for other potential imputation variables, these predictors may be important.

For all of the imputation variables, the imputation classes resulting from the CHAID analysis were not questioned, because the goal was to reduce the amount of manual processing. In the CHAID analysis, the margin for choosing one covariate over another is sometimes slim, which can result in unintuitive results, such as different imputation classes for the two parent education variables. For variables of special importance, with low levels of missing values, a manual review of the CHAID analysis results may be necessary.

The key measure for determining whether the weighted sequential hot deck imputation procedure produces acceptable results is that the before and after imputation distributions within imputation class are similar. Additionally, the overall before and after imputation distributions need to be similar as well. Distributions were considered to be similar when absolute differences were less than 2 percent where the absolute difference was calculated by subtracting the before imputation weighted percent from the after imputation weighted percent.

Tables 6 and 7 show the before and after distributions of the continuous test variables (amount of commercial loan, amount of employer aid, and hours worked per week). The weighted percentage of missing values ranged from about 2 percent (hours worked per week) to about 10 percent (amount of employer aid) (see table 5). For each of these three continuous variables, a continuous value greater than or equal to zero was imputed. Table 6 shows a breakdown of the values for those equal to zero and those greater than zero. If a case had a zero value, then the variable acted as an indicator variable. For example, if the amount of commercial loan was zero, then it was assumed that the student did not receive any funds from a commercial loan. Table 7 shows these averages (for cases with values greater than 0) before and after imputation. Regarding the commercial loan amount and employer aid amounts, students were first asked if they received any aid that did not come through the financial aid office, such as employer aid or commercial loans. If the student responded "yes," follow-up questions were asked to specify the amount of employer aid, commercial loans, etc. The imputation was limited to students who replied that they received aid, but did not specify the type or amount of these two kinds.

The before and after imputation differences in the weighted percentages for amount of commercial loan and hours worked per week were less than 1 percent. For example, for the amount of commercial loan, the weighted percentage of cases equal to zero before imputation was 87.1 percent and the weighted percentage of cases equal to zero after imputation was 86.8 percent for an absolute difference of 0.3 percentage points.

The before and after imputation distribution for amount of employer aid shows subcategory differences less than 2 percent. Table 8 shows the before and after imputation distribution for the amount of employer aid by imputation class. As shown in table 8, the absolute differences between the before and after imputation distribution for the first imputation class (dependent student who did apply for federal aid) is less than 1 percent (|7.64 – 77.1| = 0.7 and |23.6 – 22.9| = 0.7). Additionally, the average amount of employer aid was similar before and after imputation within imputation class. All other imputation class distributions show similar small differences.

Table 9 shows the before and after imputation distribution of the categorical test variables—ever vote, number of credit cards in own name, mother's education, and father's education. Table 10 shows the before and after imputation distribution for the father's education variable by imputation class. It illustrates how the weighted sequential hot deck procedure performs within imputation classes. The before and after weighted percentages within each imputation class are nearly identical.

Conclusions

When variables with low levels of missing values are imputed, many imputation procedures will produce acceptable results. The proposed method of using a "predetermined" set of predictor variables appears to be a quick, reliable, and efficient method for imputing a large number of variables with low levels of missing values. Additional investigation of procedures for integrating the CHAID analysis and imputation procedure, solely within SAS, could make the process even more efficient. An alternative procedure, which would reduce the time needed to prepare the data for the imputation procedure, would be to use the listed set of predictor variables to develop a final fixed set of imputation classes (with at least 30 respondents per class) and use this fixed set for all variables to be imputed.

Table 6. Before and after imputation distribution of continuous test (CATI) variables with low levels of missing values

	Imputed cases		Before imputation		After imputation	
Variable	**Sample size[2]**	**Weighted percent**	**Sample size[1]**	**Weighted percent**	**Sample size**	**Weighted percent**
Amount of commercial loan (NCAMNCOM) total	**470**	**100.0**	**7,720**	**100.0**	**8,190**	**100.0**
Equal to zero	370	80.8	6,390	87.1	6,760	86.8
Greater than zero	90	19.2	1,330	12.9	1,420	13.2
Amount of employer aid (NCAMNEMP) total	**560**	**100.0**	**7,630**	**100.0**	**8,190**	**100.0**
Equal to zero	320	53.5	5,760	67.4	6,070	66.1
Greater than zero	250	46.5	1,870	32.6	2,110	33.9
Hours worked per week in NPSAS year (NDHOURS) total	**740**	**100.0**	**31,990**	**100.0**	**32,710**	**100.0**
Equal to zero	170	23.0	6,980	19.4	7,150	19.5
Greater than zero	570	77.0	25,000	80.6	25,570	80.5

[1]Includes sampled undergraduate students who had a full computer-assisted telephone (CATI). The before imputation distribution sample size and weighted percents are shown for only the nonmissing cases for the respective variable.
[2]Includes sampled undergraduate students who had a full CATI interview. The after imputation distribution sample size and weighted percents are shown for both imputed and respondent cases combined for the respective variable.
NOTE: Detail may not add to total due to rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), Restricted use data file.

Table 7. Averages of continuous test variables before and after imputation

Variable	Imputed cases average	Before imputation average	After imputation average	Absolute difference	Relative percent difference
Amount of commercial loan (NCAMNCOM)	$4,800	$5,600	$5,500	$100	1.8
Amount of employer aid (NCAMNEMP)	$1,900	$1,600	$1,600	#	#
Hours worked per week in NPSAS year (NDHOURS)	30.2	31.6	31.6	#	0.1

Rounds to zero.
NOTE: Averages were calculated using sampled undergraduate students who had a full computer-assisted telephone interview (CATI).
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), Restricted use data file.

Table 8. Before and after imputation distributions for amount of employer aid

	Amount of employer aid (NCAMEMP)					
	Before imputation			After imputation		
		Weighted percent			Weighted percent	
Imputation class	Average	Zero	Greater than zero	Average	Zero	Greater than zero
Dependent students who did not apply for federal aid	$2,500	76.4	23.6	$2,700	77.1	22.9
Independent students who did not apply for federal aid	$1,400	30.1	69.9	$1,500	30.6	69.7
Federal aid applicants attending a public institution	$1,300	88.8	11.2	$1,200	89.0	11.0
Federal aid applicants attending a private institution	$2,700	89.8	10.2	$2,700	89.6	10.4

NOTE: Averages and weighted percentages were calculated using sampled undergraduate students who had a full CATI interview. The before imputation distribution sample size and weighted percents are shown for only the nonmissing cases for the respective variable. The after imputation distribution sample size and weighted percents are shown for both imputed and respondent cases combined. Detail may not add to total due to rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), Restricted use data file.

Table 9. Before and after imputation distribution of categorical test (CATI) variables with low levels of missing values

Variable	Imputed cases		Before imputation		After imputation	
	Sample size	Weighted percent	Sample size[1]	Weighted percent	Sample size[2]	Weighted percent
Ever vote in 1999–2000 (NBEVRVT) total	**180**	**100.0**	**30,720**	**100.0**	**30,910**	**100.0**
Yes	120	59.8	20,370	63.7	20,490	63.7
No	60	40.2	10,350	36.3	10,420	36.3
Number of credit cards in own name (NDNUMCRD) total	**1,110**	**100.0**	**31,670**	**100.0**	**32,780**	**100.0**
0	300	25.2	8,490	29.4	8,790	29.3
1	500	45.4	13,900	42.5	14,390	42.6
2	320	29.4	9,280	28.1	9,600	28.2
Mother's education (NBMOMED) total	**1,390**	**100.0**	**31,350**	**100.0**	**32,740**	**100.0**
Less than high school	190	13.2	3,120	11.6	3,310	11.7
High school diploma or equivalent	580	40.9	11,780	39.7	12,360	39.7
Some college\associate's degree[3]	300	22.9	7,130	22.5	7,430	22.5
Completed bachelor's degree	210	15.1	6,060	17.7	6,270	17.6
Completed master's degree or beyond	110	7.9	3,260	8.6	3,370	8.5
Father's education (NBDADED) total	**2,340**	**100.0**	**30,240**	**100.0**	**32,580**	**100.0**
Less than high school	350	14.9	3,390	12.8	3,740	13.0
High school diploma or equivalent	830	34.8	10,030	34.9	10,860	34.9
Some college\associate's degree[3]	440	17.3	5,410	17.9	5,850	17.8
Completed bachelor's degree	430	19.5	6,370	20.4	6,810	20.4
Completed master's degree or beyond	290	13.4	5,040	14.0	5,330	14.0

[1]Includes sampled undergraduate students who had a full computer-assisted telephone interview (CATI).The before imputation distribution sample size and weighted percents are shown for only the nonmissing cases for the respective variable.
[2]Includes sampled undergraduate students who had a full CATI interview. The after imputation distribution sample size and weighted percents are shown for both imputed and respondent cases combined for the respective variable.
[3]Some college includes vocational/technical training and attending college but not receiving a bachelor's degree.
NOTE: Detail may not add to total due to rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), Restricted use data file.

Table 10. Before and after imputation distribution (weighted percents) of father's education, by imputation class

Imputation class	Total percent	Less than high school	High school diploma or equivalent	Some college	Completed bachelor's degree	Completed master's degree or beyond
			Before imputation			
Dependents who did not apply for federal aid	100.0	4.7	26.0	18.1	28.3	23.0
Dependents who applied for federal aid	100.0	6.5	36.5	20.8	22.0	14.2
Independents who had no dependents	100.0	15.9	36.0	16.7	18.9	12.4
Independents with dependents	100.0	23.9	40.0	15.4	13.3	7.4
			After imputation			
Dependents who did not apply for federal aid	100.0	4.7	25.8	17.7	28.6	23.2
Dependents who applied for federal aid	100.0	6.5	36.5	20.9	22.0	14.1
Independents who had no dependents	100.0	15.9	35.9	16.8	18.9	12.5
Independents with dependents	100.0	24.0	39.9	15.6	13.1	7.4

NOTE: Weighted percentages were calculated using sampled undergraduate students who had a full CATI interview. The before imputation distribution sample size and weighted percents are shown for only the nonmissing cases for the respective variable. The after imputation distribution sample size and weighted percents are shown for both imputed and respondent cases combined. Detail may not add to total due to rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), Restricted use data file.

Chapter 4

Method for Imputing Price of Attendance

Statement of Objective

The price of attendance (student budget) is an important variable in the NPSAS study that is available in the institutional survey (CADE) and is reported for those students who applied for financial aid (about one-half of the sample cases). Prior imputation methods have been based on calculating average budgets for categories of students who applied for aid (using combinations of student dependency status and student housing) and attributing these averages to students with the same characteristics who did not apply for aid (and therefore had no student budget). This method of calculating averages, which consists of subsetting the data and applying consistency checks, is very time consuming and sometimes produces inconsistent results. The third study objective aimed to investigate a quicker and more efficient method for imputing the price of attendance.

The student budget is the sum of tuition and nontuition expenses. The nontuition costs are books, room and board, transportation and personal expenses, and sometimes a computer allowance. Tuition amounts are highly dependent on institution control (public, private not-for-profit, private for-profit) and the level of the institution (4-year, 2-year, less-than-2-year) and have a wide range (from $100 at some community colleges to over $30,000 at some private colleges). Nontuition costs, on the other hand, have a relatively narrow range. Nontuition costs reported in NPSAS:2000 typically ranged from $6,000 to $10,000 for full-time undergraduates and were related to student dependency status (dependent versus independent) and student housing (living on campus, living off campus, or living at home with parents).

Methodology

To test the imputation procedure, student budgets were only imputed for undergraduates who were enrolled full time for the full academic year at a single institution during the NPSAS year and who did not become graduate students during the year (about 17,000 cases). Tuition amounts were reported separately by the institutions in CADE and had very few missing values. Therefore, the imputation procedure tested assumed that tuition was not missing for any students and only imputed the nontuition portion of the price of attendance. Because institutions typically do not report the nontuition student budget components separately, the total of all nontuition costs was imputed.

Weighted sequential hot deck imputation was used to impute the nontuition costs. A CHAID analysis was performed using the main predictors that were used in the previous NPSAS:2000 imputation process. These predictors were institution (INSTID); institution state (INSSTATE); institution control (public, private not-for-profit, private for-profit) (CONTROL); institution level (4-year, 2-year, less-than-2-year) (LEVEL); student housing (LOCALRES); and student dependency (DEPEND2). In addition, predictors included total amount of financial aid received (TOTAID), federal need-based aid received (FEDNEED), the expected family contribution

(EFC4), the total amount of need-based financial aid received (EFCAID1), and tuition (TUITION2). Appendix D shows the imputation classes from the CHAID analysis for the nontuition costs variable (SBNONTUN).

After imputation of the nontuition costs, the total price of attendance for full-time, full-year students (BUDGETFT) was calculated as the sum of tuition and nontuition (SBNONTUN) expenses. The student budget has the following constraints: (a) the budget must be greater than or equal to the total amount of financial aid received, and (b) if the student received need-based federal aid (FEDNEED>0), then the budget must be greater than or equal to the sum of the federal expected family contribution and the amount of need-based financial aid received.

Results

Table 11 below shows the distribution of nontuition costs and the price of attendance variables before and after imputation and by the prior and the proposed methods. The criteria for judging acceptable results were similar to those used for the procedure discussed in chapter 2. Relative percent differences between the prior and proposed imputation distributions were calculated using the formula in chapter 2 and differences less than 10 percent were considered acceptable. All relative percent differences were within the acceptable range.

The average nontuition costs after imputation with the new methodology were about 5 percent lower than the after imputation results using the prior methodology. This was not unexpected, because the results from the prior method were based on two steps. First, a preliminary estimate was imputed, based on the average amounts for various categories of students within an institution or similar institutions. Then the preliminary averages for individual cases were adjusted in about 5 percent of the cases to meet the two constraints discussed above, which always required the student budget to be increased. Since it was assumed that the tuition component was correct, the adjustment had to increase the nontuition component, and the method biased the results upward.

The results from the new methodology were checked to see how many cases fell outside the bounds of the two constraints. The first constraint was that the total student budget had to be greater than or equal to the total amount of financial aid received. There were 297 cases (1.8 percent) that did not meet this constraint. The second constraint was that for students who received federal need-based aid, the total budget had to be greater than or equal to the sum of the expected family contribution and the total amount of need-based aid received. There were 102 cases (0.6 percent) that did not meet this second constraint.

Conclusions

The past student budget imputation procedures based on averages typically resulted in about 5 percent of the cases that did not meet the constraints, and therefore required an adjustment that introduced an upward bias in the final imputed amounts. The proposed method produced a relatively small percentage of cases (about 2 percent) that did not meet the constraints, so a similar postimputation adjustment would introduce less bias. The proposed method also has the advantage of being more efficient in processing a large number of cases faster than the prior method.

Table 11. Before and after imputation distributions of price of attendance variables, by method

Variable	After imputation mean (prior method)	Before imputation mean	Imputed cases mean	After imputation mean (proposed method)	Absolute difference	Relative percent difference
Nontuition costs (SUBNONTUN) total	**$8,300**	**$8,000**	**$7,700**	**$7,900**	$400	**–4.9**
Public 4-year	8,400	8,100	7,800	8,000	400	–4.9
Public 2-year	7,500	7,300	7,000	7,100	400	–5.5
Private not-for-profit 4-year	8,600	8,100	8,100	8,100	500	–6.0
Private for-profit	9,500	9,400	9,000	9,300	200	–2.1
Dependent	7,995	7,700	7,500	7,600	400	–5.1
Independent with dependents	9,100	8,800	8,400	8,600	500	–5.6
Independent with no dependents	9,000	8,900	8,300	8,700	300	–3.4
On-campus	8,400	8,100	8,100	8,100	300	–3.6
Off-campus	9,000	8,700	8,400	8,600	400	–4.5
Living with parents	7,000	6,500	6,300	6,400	600	–9.0
Price of attendance for full-time, full-year student (BUDGETFT) total	**$14,800**	**$14,900**	**$13,500**	**$14,300**	500	**–3.4**
Public 4-year	12,600	12,200	12,200	12,200	400	–3.2
Public 2-year	9,100	8,900	8,500	8,700	400	–4.5
Private not-for-profit 4-year	23,700	22,500	25,100	23,400	300	–1.3
Private for-profit	18,500	18,500	17,800	18,300	200	–1.1
Dependent	15,100	15,500	13,600	14,600	500	–3.4
Independent with dependents	14,000	14,100	12,600	13,500	500	–3.6
Independent with no dependents	13,500	12,800	13,300	13,000	500	–3.8
On-campus	18,400	18,400	17,300	18,000	400	–2.2
Off-campus	14,400	13,900	13,900	13,900	500	–3.5
Living with parents	11,400	11,200	10,000	10,600	800	–7.3

NOTE: Includes sampled undergraduates who were enrolled full time for the full academic year at a single institution during the NPSAS year and who did not become graduate students during the year. Detail may not add to total due to rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), Restricted use data file.

This page intentionally left blank.

Chapter 5

Estimation of the Relative Magnitude of the Imputation Component of Variance

Statement of Objective

Virtually all imputation procedures include stochastic components. Consequently, with imputed data, use of conventional methods of variance estimation can fail to reflect the added variability due to imputation. Moreover, with most imputation procedures, missing values are replaced by a subset of observed values. This induced homogeneity, which results from repeated usage of certain observed values, can bring about further deflation of variance estimates. There is a large body of research on this topic, including those studies conducted by Little and Rubin (1987), Sarndal (1992), Rao and Shao (1992), Rao and Sitter (1995), Lee, Rancourt, and Sarndal (1995), and Schafer (1997).

Because of the creative techniques that are typically used to impute missing data, there is no single method for estimation of variances with imputed data that would be suitable for all survey situations. This is particularly true when data are obtained through complex survey designs, where variance estimation is compounded by multifaceted weighting adjustments. To address the study's fourth objective, this section presents a heuristic approach for assessing some of the variance inflations that occur because of imputing missing data.

Methodology

To achieve the goals of this objective, using the weighted hot deck imputation methodology described earlier, 10 complete "decks" of data were created. That is, the same imputation process was used with 10 different seeds for generation of random numbers. Subsequently, using SUDAAN, 10 separate point estimates of means and their corresponding standard errors were calculated for the following five continuous variables:

- Amount of commercial loan (NCAMNCOM);
- Amount of employer aid (NCAMNEMP);
- Balance due on all credit cards (NDCRDBAL);
- Hours worked per week in NPSAS year (NDHOURS); and
- Earnings from work while enrolled (WKINC).

Next, following the methodology suggested by Schafer (1997, p. 109) for each estimate of mean, two components of variance were calculated: one to reflect the within-imputation variance, and another to reflect the between-imputation variance. The former was simply the average of variance estimates for the given parameter obtained from each of the m = 10 decks, while the latter was the sample variance of the resulting 10 point estimates. That is, for a parameter θ the within- and between-imputation variance components of its estimate $\hat{\theta}$ were given by:

$$W(\hat{\theta}) = \frac{1}{m}\sum_{i=1}^{m}\hat{V}_i(\hat{\theta}) \text{ and } B(\hat{\theta}) = \frac{1}{m-1}\sum_{i=1}^{m}\left(\hat{\theta}_i - \bar{\hat{\theta}}\right)^2$$

resulting in a total variance estimate that attempts to reflect the inflation due to imputation, defined as:

$$V(\hat{\theta}) = W(\hat{\theta}) + \left(1 + \frac{1}{m}\right)B(\hat{\theta}).$$

Results

To achieve a more comprehensive understanding of the components of variance inflation, prior to estimating the mean for each of the selected variables, the above calculations were first carried out for the percent of cases that had a nonzero value (imputed or observed) for each variable. The results are summarized in table 12 below, where percent imputed excludes inapplicable cases, coded in the data as –3.

Table 12. Variance components for the percentages with positive values for the five continuous variables

Variable	Applicable cases	Percent imputed	$\hat{\theta}$	$W(\hat{\theta})$	$B(\hat{\theta})$	$V(\hat{\theta})$	$\frac{V(\hat{\theta})}{W(\hat{\theta})}-1$
Hours worked per week in NPSAS year	32,710	2.2	80.56%	0.1590	0.0018	0.1609	1.22%
Amount of commercial loan	8,190	5.7	13.07%	0.2430	0.0139	0.2583	6.30%
Amount of employer aid	8,190	6.9	34.13%	0.7120	0.0240	0.7384	3.71%
Earnings from work while enrolled	32,780	11.5	78.37%	0.1590	0.0053	0.1649	3.69%
Balance due on all credit cards	12,390	16.6	98.10%	0.0330	0.0061	0.0397	20.28%

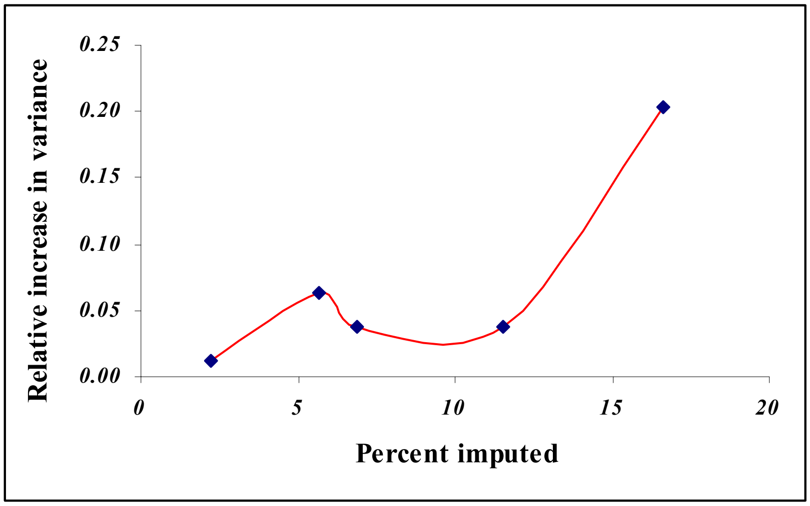
NOTE: Includes sampled undergraduate students who had a full computer-assisted telephone interview (CATI).
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), Restricted use data file.

To illustrate the calculations in the above table, we will consider those carried out for commercial loan. The 10 estimates of the percent cases with nonzero commercial loan have a mean of 13.07 percent and a variance of 0.014, while the corresponding estimates of variance of this parameter have an average of 0.243. In this case, the estimates of within- and between-imputation variance components would be 0.243 and 0.014, respectively, resulting in an estimated relative increase in variance due to imputation given by:

$$\frac{W(\hat{\theta})}{V(\hat{\theta})} - 1 = \frac{0.243 + \left(1 + \frac{1}{10}\right)0.014}{0.243} - 1 = 6.3\%$$

The following figure is a pictorial representation of the results from table 12, illustrating the nature of the relation between percent imputed data and variance inflation due to imputation.

Figure 1. Relative increase in variances as a function of percent imputed data for the percentages with positive values for the five continuous variables



SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 1999-2000 (NPSAS:2000), Restricted use data file.

Table 13 provides a summary of relevant statistics for estimates of mean for the above variables based on the 10 imputed datasets. Note that in this table percent imputed for each variable is an average over the 10 decks and excludes both zero and inapplicable cases.

Moreover, the same set of measures as those summarized in table 12 was obtained for estimates of percent of students based on their parents' education, the results of which are summarized in table 14 and figure 2. Note that for these analyses the original nine categories of parents' education were collapsed into the following six categories:

1. Less than high school;
2. High school or equivalent;
3. Some college;
4. Associate's degree;
5. Bachelor's degree; and
6. Master's degree and beyond.

Analogous to table 12, table 14 provides a summary of the within- and between-imputation variance components for estimates of percents for the above variables based on the 10 imputed datasets.

Table 13. Variance components for the amounts for the five continuous variables

Variable	Positive cases	Percent imputed	$\hat{\theta}$	$W(\hat{\theta})$	$B(\hat{\theta})$	$V(\hat{\theta})$	$\frac{V(\hat{\theta})}{W(\hat{\theta})}-1$
Hours worked per week in NPSAS year	25,570	2.9	31.61	0.023	0.0003	0.02	1.4%
Earnings from work while enrolled	24,950	15.1	13,532.95	30,443.625	1,525.762	32,121.96	5.5%
Balance due on all credit cards	12,140	17.0	3,067.86	3,400.576	475.022	3,923.10	15.4%
Amount of employer aid	2,110	26.5	1,577.95	7,707.213	1,195.361	9,022.11	17.1%
Amount of commercial loan	1,420	32.9	5,618.83	82,822.468	4,027.679	87,252.91	5.3%

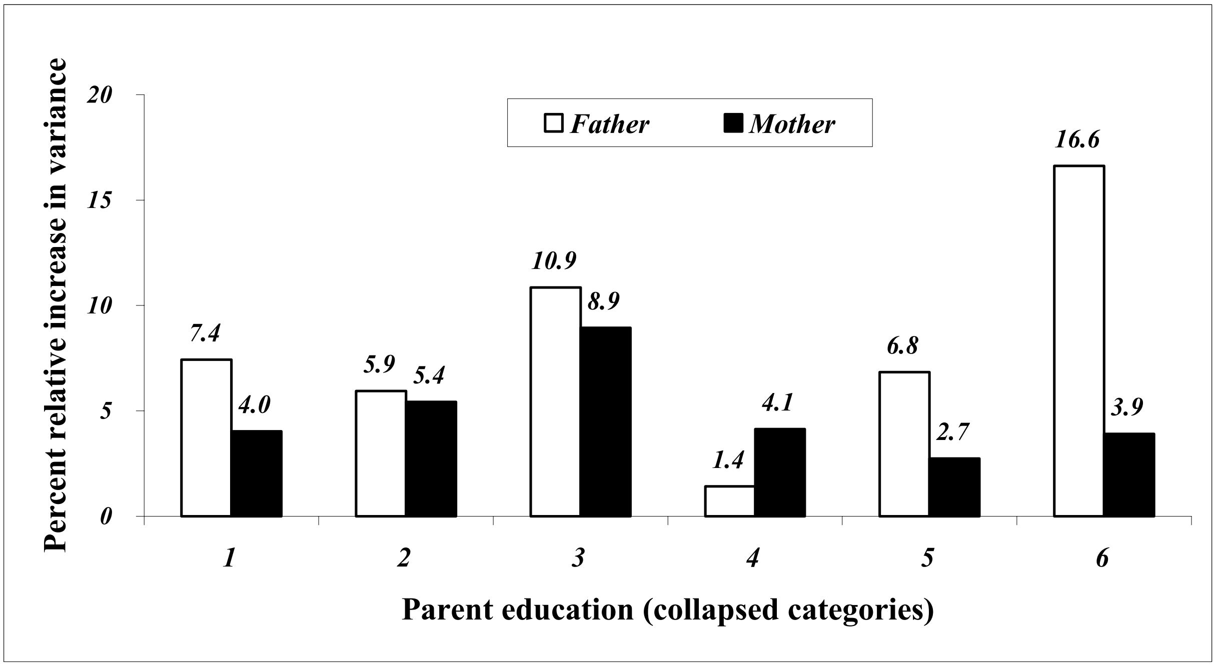
NOTE: Includes sampled undergraduate students who had a full computer-assisted telephone interview (CATI).
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), Restricted use data file.

Table 14. Variance components for estimates of percent students, by parent education

Variable	Applicable cases	Percent imputed	$\hat{\theta}$	$W(\hat{\theta})$	$B(\hat{\theta})$	$V(\hat{\theta})$	$\frac{V(\hat{\theta})}{W(\hat{\theta})}-1$
Father's education							
Less than high school	3,740	9.5	12.94%	0.128	0.009	0.138	7.43%
High school or equivalent	10,870	7.7	35.00%	0.200	0.011	0.212	5.95%
Some college	2,880	6.5	8.67%	0.054	0.005	0.060	10.86%
Associate's degree	2,930	7.2	9.16%	0.061	0.001	0.062	1.41%
Bachelor's degree	6,830	6.7	20.29%	0.139	0.009	0.149	6.84%
Master's degree and beyond	5,330	5.4	13.94%	0.083	0.013	0.097	16.62%
Mother's education							
Less than high school	3,310	5.7	11.72%	0.134	0.005	0.139	4.02%
High school or equivalent	12,340	4.6	39.69%	0.133	0.007	0.140	5.43%
Some college	3,310	3.9	10.17%	0.052	0.004	0.057	8.94%
Associate's degree	4,100	3.8	12.23%	0.072	0.003	0.075	4.13%
Bachelor's degree	6,290	3.7	17.65%	0.109	0.003	0.112	2.73%
Master's degree and beyond	3,380	3.5	8.55%	0.045	0.002	0.047	3.91%

NOTE: Includes sampled undergraduate students who had a full computer-assisted telephone interview (CATI).
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), Restricted use data file.

Figure 2. Percent relative increase in variances due to imputation for estimates of percent students, by parent education



NOTE: 1=less than high school; 2=high school or equivalent; 3=some college; 4=associate's degree; 5=bachelor's degree; 6=master's degree and beyond. Includes sampled undergraduate students who had a full CATI interview.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), Restricted use data file.

Conclusions

As stated earlier, there is a growing body of research suggesting that variance estimation in the presence of imputed data needs to reflect the added variability that will result from the stochastic nature of imputation. While scientific debates continue regarding a general method for defining and capturing this variance inflation, there are empirical and intuitive justifications for supporting the hypothesis that the resulting increase in variances is a function of the severity of missing data.

As seen in table 12, this research supports the above assertion. Although the number of estimates involved in this analysis is small, there nonetheless appears to be a positive linear relation between the unweighted percent of missing data and the relative increase in variances. Moreover, specific to the NPSAS:2000 data, the results suggest that the magnitude of the variance inflation due to imputation can be sizable.[1] Consequently, it may be imprudent to completely ignore variance inflation due to imputation for NPSAS data.

[1] Since there is not a singularly accepted method for measuring variance inflation due to imputation, it would be reassuring to use other methods to estimate this inflation. Some of the estimates obtained based on the employed method are surprisingly high.

On the other hand, any significant change in the variance estimation methodology for a survey of this complexity can entail major operational ramifications, particularly in connection with the public release data. A change of this nature also can introduce complications for analyses that involve comparisons of results across different administrations of the survey. Thus, it would be wise to conduct more comprehensive research before specific changes are implemented in variance estimation for NPSAS. Other methods of measuring variance inflation due to imputation, besides that examined here, should be investigated. An excellent opportunity would be to conduct similar research using data from the National Study of Postsecondary Faculty: 1999 (NSOPF:99). The NPSAS and NSOPF studies have many common features, and they are now, for the first time, being administered as a joint survey. Such research can help not only re-examine the findings of this study, but also to serve as a preparation for the imputation of NSOPF:2004 data as well.

Chapter 6

Improvements in Data Processing

This chapter summarizes the results of the four different study objectives and suggests improvements and recommendations for implementing the new imputation methodology.

Objective 1: Imputation procedure for resolving discrepancies between data sources.

Improvement: Discrepancies between data sources can result from both valid changes in status that take place over time and reporting errors. Both the prior and proposed methods produce acceptable results and there are no real criteria for judging which of the two methods is "better." To improve the speed of the data editing and imputation process, inconsistencies that are easily resolved should be resolved with logical imputations. The remaining inconsistencies should be set to missing values and imputed using stochastic imputation.

Objective 2: Imputation procedure for variables with low levels of missing values.

Improvement: The proposed method of using a "predetermined set" of predictor variables to form imputation classes for the weighted sequential hot deck imputation procedure has been determined to be a reliable and efficient method. However, more investigation is needed to determine whether a "predetermined set" of *imputation classes* can be used for all variables, which would reduce the time needed for the imputation process even more. Additionally, a closer examination of the list of potential predictor variables (table 5) is needed to determine if any variables should be added or deleted. One issue to consider before finalizing the list is that several variables in the list require imputation first. The current list contains two sets of essentially redundant variables of this sort: (1) NBUSBORN and CITIZEN2 and (2) INCPS and FEDAPP. We probably should use only one of the two variables in each case. Finally, we suggest that similar research be conducted using other survey data such as the NSOPF data.

Objective 3: Imputation procedure for price of attendance.

Improvement: The past student budget imputation procedures required a method of calculating averages and then applying consistency checks. The proposed method appears to be more efficient in processing a large number of cases faster than the prior method. The nontuition costs for the price of attendance variable should be imputed using weighted sequential hot deck imputation and performing a CHAID analysis on the NPSAS:2004 data, using the same variables and possible additional variables to determine imputation classes.

Objective 4: Estimation of variance inflation due to imputation.

Improvement: Previous studies suggest that variance estimation in the presence of imputed data has to reflect the added variability that will result from the random nature of imputation. While there is not a singularly accepted method for defining and measuring variance inflation due to imputation, the employed methodology in this research supports the fundamental concept that the resulting increase in variances is a function of the severity of missing data. Moreover, the corresponding findings suggest that the magnitude of this inflation can be sizeable (i.e., 10 percent or more).

Because changing the variance estimation methodology can entail major operational ramifications in connection with the public release data, it is advisable to conduct more comprehensive research before specific changes are implemented in variance estimation for NPSAS.

References

Cox, B.G. (1980). The Weighted Sequential Hot Deck Imputation Procedure. In *Proceedings of the Survey Research Methods Section* (pp. 721–726). The American Statistical Association.

Iannacchione, V.G. (1982). Weighted Sequential Hot Deck Imputation Macros. *In Proceedings of the Seventh Annual SAS User's Group International Conference* (pp.759–763). Cary, NC: SAS Institute, Inc.

Little, R.J.A. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. New York : Wiley.

Rao, J.N.K., and Shao, J. (1992). Jackknife Variance Estimation with Survey Data under Hot Deck Imputation. *Biometrika, 79*: 811–822.

Sarndal, C. E. (1992). Methods for Estimating the Precision of Survey Estimates When Imputation Has Been Used, *Survey Methodology*, *19*: 241–252.

Rao, J.N.K., and Sitter, R. R. (1995). Variance Estimation Under Two-Phase Sampling with Applications to Imputation for Missing Data, *Biometrika*, *63*: 581–590.

Lee, H., Rancourt, E., and Sarndal, C.E. (1995). Variance Estimation in the Presence of Imputed Data for the Generalized Estimation System. In *Proceedings of the Section on Survey Research Methods* (pp. 384–389). The American Statistical Association.

Schafer, J.L. (1997). *Analysis of Incomplete Multivariate Data*. Monographs on Statistics and Applied Probability, vol. 72. London: Chapman & Hall/CRC Press.

This page intentionally left blank.

Appendix A

Imputation Classes for Test Variables as from the Chi-squared Automatic Detection (CHAID) Analysis

Table A–1. Imputation classes for test variables as from Chi-squared automatic interaction detection (CHAID) analysis

Variable	Imputation classes
Student citizenship (CITIZEN2)	Not born in U.S., not in CPS[1], English as primary language
	Not born in U.S., not in CPS, non-English as primary language
	Not born in U.S., in CPS, attending 4-year institution
	Not born in U.S., in CPS, attending 2-year or less-than-2-year institution
	Born in U.S., not in CPS, attending 4-year or 2-year institution
	Born in U.S., not in CPS, attending less-than-2-year institution
	Born in U.S., in CPS, resident alien, citizen foreign born self or parent
	Born in U.S., in CPS, citizen
Student marital status (SMARITAL)	Not in CPS, 23 years old or younger, did not apply for federal aid in 1999–2000
	Not in CPS, 23 years old or younger, applied for federal aid in 1999–2000
	Not in CPS, between 24 and 29 years old, working students
	Not in CPS, between 24 and 29 years old, employees who studied or those who did not work
	Not in CPS, over 29 years old, enrolled in a 4-year institution
	Not in CPS, over 29 years old , enrolled in 2- or less–than-2-year institution
	In CPS, between 24 and 29 years old , males who did not work while enrolled
	In CPS, between 24 and 29 years old, females who did not work while enrolled
	In CPS, between 24 and 29 years old, worked part time while enrolled
	In CPS, between 24 and 29 years old, worked full time while enrolled
Student had any dependents (ANYDEP)	Not in CPS, single, did not apply for federal aid
	Not in CPS, single, applied for federal aid
	Not in CPS, married or separated, did not apply for federal aid
	Not in CPS, married or separated, applied for federal aid
	In CPS, single males
	In CPS, married males
	In CPS, separated males
	In CPS, females who did not work while enrolled
	In CPS, females who worked while enrolled
Number of dependents (NDEPEND)	Not in CPS, married, did not apply for federal aid
	Not in CPS, married, applied for federal aid
	In CPS, married, did not apply for federal aid
	In CPS, married, applied for federal aid
	Not in CPS, married or separated, did not apply for federal aid
	Not in CPS, married or separated, applied for federal aid
	In CPS, married or separated, males
	In CPS, married or separated females

[1] CPS=Central Processing System.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), Restricted use data file.

Appendix B
After Imputation Distribution of Student Marital Status by Imputation Class and Method

Table B–1. Weighted relative percent differences between proposed and prior Imputation Procedures, by imputation class for student marital status

Imputation classes	Sample size[1]	Weighted percent difference		
		Single	Married	Separated
Total	**32,780**	**2.1**	**4.6**	**66.3**
Not in CPS[2], 23 years or younger, did not apply for federal aid	7,200	0.7	–22.5	72.7
Not in CPS, 23 years or younger, applied for federal aid	300	3.9	–30.4	0.0
Not in CPS, between 24 and 29 years old, working students	960	3.0	–12.7	16.7
Not in CPS, between 24 and 29 years old, employees who studied or those who did not work	1,280	-2.9	4.8	0.9
Not in CPS, over 29 years old, enrolled in a 4-year institution	2,150	6.0	–2.7	–21.1
Not in CPS, over 29 years, enrolled in a 2- or less-than-2-year institution	1,890	1.7	–2.3	33.2
In CPS, between 24 and 29 years old, males who did not work while enrolled	1,530	1.7	–5.7	–145.6
In CPS, between 24 and 29 years old, females who did not work while enrolled	2,630	4.5	–6.8	–120.1
In CPS, between 24 and 29 years old, worked part time while enrolled	10,380	2.3	–12.8	–209.3
In CPS, between 24 and 29 years old, worked full time while enrolled	4,480	4.7	–5.4	–232.0

[1]Includes sampled undergraduate students who had a full computer-assisted telephone interview (CATI) and who were U.S. citizens.
[2]CPS=Central Processing System.
NOTE: Detail may not add to total due to rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), Restricted use data file.

Appendix C

Imputation Classes for Test Variables with Low Levels of Missing Values

Table C–1. Imputation classes for test variables with low levels of missing values

Variable	Imputation classes
Father's education (NBDADED)	Dependents in 1999–2000 who did not apply for federal aid Dependents in 1999–2000 who applied for federal aid Independents in 1999 who had no dependents Independents with dependents
Ever vote (NBEVRVT)	Dependents, living with parents or on campus Dependents, living off campus Independents, single or separated Independents, married
Mother's education (NBMOMED)	Dependents, non-Hispanic Dependents, Hispanic Independents, non-Hispanic Independents, Hispanic
Amount of commercial loan (NCAMNCOM)	Did not apply for federal aid attending public institution Did not apply for federal aid attending private institution Federal aid applicants in 1999–2000 who lived on campus Federal aid applicants in 1999–2000 who lived off campus
Amount of employer aid (NCAMNEMP)	Dependents who did not apply for federal aid Independents who did not apply for federal aid Federal aid applicants attending public institution Federal aid applicants attending private institution
Number of hours worked for week in NPSAS year (NDHOURS)	Dependents who were either citizens or resident aliens Dependents who were foreign students Independents who did not apply for federal aid Independents who applied for federal aid
Number of credit cards in own name (NDNUMCRD)	Those 15–19 years old attending school in outlying areas (PR) Those over 19 and at most 48 years old attending school in PR Dependents attending 4-year institution within the U.S. Independents attending 4-year institution within the U.S. Those 15–19 years old attending 2-year institution within the U.S. Those over 19 and at most 84 years old attending 2-year institution in U.S. White students attending less-than-2-year institution in the U.S Black, Hispanic, Asian, and Indian students attending less-than-2-year institution in the U.S.
Sort variables for all imputations	
OBE[1] region code 1999–2000 (OBEREG)	
Institution state (INSSTATE)	
Degree of urbanization (LOCALE)	
Control 1999–2000 (CONTROL)	
NPSAS sample institution level 1999–2000 (LEVEL)	
NPSAS sample institution identification (INSTID)	
Student's citizenship in 1999–2000 (CITIZEN2)	
Race category (RACECAT)	
Gender 1999–2000 (GENDER)	
Age as of 12/31/99 1999–2000 (AGE)	
Attendance intensity in fall 1999–2000 (ATTEND)	
Dependency status 1999–2000 (DEPEND)	

[1]OBE region=Bureau of Economic Analysis Region Code.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), Restricted use data file.

Appendix D
Imputation Classes for Nontuition Cost For Price of Attendance

Table D–1. Imputation classes for nontuition cost for price of attendance

Dependent or unknown dependence status & 4-year institution & living on campus
Dependent or unknown dependence status & 4-year institution & living off campus
Dependent or unknown dependence status & 4-year institution & living with parents
Dependent or unknown dependence status & 2-year or less-than-2-year institution & not living with parents
Dependent or unknown dependence status & 2-year or less-than-2-year institution & living with parents
Independent & total aid = $0 – $7,359 & not living with parents
Independent & total aid = $0 – $7,359 & living with parents
Independent & total aid = $7,360 – $10,766 & no expected family contribution
Independent & total aid = $7,360 – $10,766 & expected family contribution = $1 – $5,184
Independent & total aid = $7,360 – $10,766 & expected family contribution > $5,184
Independent & total aid = $10,767 – $37,518 & tuition and fees ≤ $3,955
Independent & total aid = $10,767 – $37,518 & tuition and fees = $3,956 – $4,981
Independent & total aid = $10,767 – $37,518 & tuition and fees = $4,982 – $14,292
Independent & total aid = $10,767 – $37,518 & tuition and fees = $14,293 – $17,945
Independent & total aid = $10,767 – $37,518 & tuition and fees > $17,945

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), Restricted use data file.

Listing of NCES Working Papers to Date

Working papers can be downloaded as .pdf files from the NCES Electronic Catalog (http://nces.ed.gov/pubsearch/). You can also contact Sheilah Jupiter at (202) 502–7363 (sheilah.jupiter@ed.gov) if you are interested in any of the following papers.

Listing of NCES Working Papers by Program Area

No.	Title	NCES contact
Baccalaureate and Beyond (B&B)		
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
2001–15	Baccalaureate and Beyond Longitudinal Study: 2000/01 Follow-Up Field Test Methodology Report	Andrew G. Malizio
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
Beginning Postsecondary Students (BPS) Longitudinal Study		
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D'Amico
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
1999–15	Projected Postsecondary Outcomes of 1992 High School Graduates	Aurora D'Amico
2001–04	Beginning Postsecondary Students Longitudinal Study: 1996–2001 (BPS:1996/2001) Field Test Methodology Report	Paula Knepper
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
Common Core of Data (CCD)		
95–12	Rural Education Data User's Guide	Samuel Peng
96–19	Assessment and Analysis of School-Level Expenditures	William J. Fowler, Jr.
97–15	Customer Service Survey: Common Core of Data Coordinators	Lee Hoffman
97–43	Measuring Inflation in Public School Costs	William J. Fowler, Jr.
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
1999–03	Evaluation of the 1996–97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle	Beth Young
2000–12	Coverage Evaluation of the 1994–95 Common Core of Data: Public Elementary/Secondary School Universe Survey	Beth Young
2000–13	Non-professional Staff in the Schools and Staffing Survey (SASS) and Common Core of Data (CCD)	Kerry Gruber
2002–02	School Locale Codes 1987 - 2000	Frank Johnson
Data Development		
2000–16a	Lifelong Learning NCES Task Force: Final Report Volume I	Lisa Hudson
2000–16b	Lifelong Learning NCES Task Force: Final Report Volume II	Lisa Hudson
Decennial Census School District Project		
95–12	Rural Education Data User's Guide	Samuel Peng
96–04	Census Mapping Project/School District Data Book	Tai Phan
98–07	Decennial Census School District Project Planning Report	Tai Phan
Early Childhood Longitudinal Study (ECLS)		
96–08	How Accurate are Teacher Judgments of Students' Academic Performance?	Jerry West
96–18	Assessment of Social Competence, Adaptive Behaviors, and Approaches to Learning with Young Children	Jerry West
97–24	Formulating a Design for the ECLS: A Review of Longitudinal Studies	Jerry West
97–36	Measuring the Quality of Program Environments in Head Start and Other Early Childhood Programs: A Review and Recommendations for Future Research	Jerry West
1999–01	A Birth Cohort Study: Conceptual and Design Considerations and Rationale	Jerry West
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
2001–02	Measuring Father Involvement in Young Children's Lives: Recommendations for a Fatherhood Module for the ECLS-B	Jerry West
2001–03	Measures of Socio-Emotional Development in Middle Childhood	Elvira Hausken

No.	Title	NCES contact
2001–06	Papers from the Early Childhood Longitudinal Studies Program: Presented at the 2001 AERA and SRCD Meetings	Jerry West
2002-05	Early Childhood Longitudinal Study-Kindergarten Class of 1998–99 (ECLS–K), Psychometric Report for Kindergarten Through First Grade	Elvira Hausken
Education Finance Statistics Center (EDFIN)		
94–05	Cost-of-Education Differentials Across the States	William J. Fowler, Jr.
96–19	Assessment and Analysis of School-Level Expenditures	William J. Fowler, Jr.
97–43	Measuring Inflation in Public School Costs	William J. Fowler, Jr.
98–04	Geographic Variations in Public Schools' Costs	William J. Fowler, Jr.
1999–16	Measuring Resources in Education: From Accounting to the Resource Cost Model Approach	William J. Fowler, Jr.
Education Longitudinal Study: 2002 (ELS:2002)		
2003-03	Education Longitudinal Study: 2002 (ELS: 2002) Field Test Report	Jeffrey Owings
High School and Beyond (HS&B)		
95–12	Rural Education Data User's Guide	Samuel Peng
1999–05	Procedures Guide for Transcript Studies	**Dawn Nelson**
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
HS Transcript Studies		
1999–05	Procedures Guide for Transcript Studies	**Dawn Nelson**
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson
2003–01	Mathematics, Foreign Language, and Science Coursetaking and the NELS:88 Transcript Data	Jeffrey Owings
2003–02	English Coursetaking and the NELS:88 Transcript Data	Jeffrey Owings
International Adult Literacy Survey (IALS)		
97–33	Adult Literacy: An International Perspective	Marilyn Binkley
Integrated Postsecondary Education Data System (IPEDS)		
97–27	Pilot Test of IPEDS Finance Survey	Peter Stowe
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
2000–14	IPEDS Finance Data Comparisons Under the 1997 Financial Accounting Standards for Private, Not-for-Profit Institutes: A Concept Paper	Peter Stowe
National Assessment of Adult Literacy (NAAL)		
98–17	Developing the National Assessment of Adult Literacy: Recommendations from Stakeholders	Sheida White
1999–09a	1992 National Adult Literacy Survey: An Overview	Alex Sedlacek
1999–09b	1992 National Adult Literacy Survey: Sample Design	Alex Sedlacek
1999–09c	1992 National Adult Literacy Survey: Weighting and Population Estimates	Alex Sedlacek
1999–09d	1992 National Adult Literacy Survey: Development of the Survey Instruments	Alex Sedlacek
1999–09e	1992 National Adult Literacy Survey: Scaling and Proficiency Estimates	Alex Sedlacek
1999–09f	1992 National Adult Literacy Survey: Interpreting the Adult Literacy Scales and Literacy Levels	Alex Sedlacek
1999–09g	1992 National Adult Literacy Survey: Literacy Levels and the Response Probability Convention	Alex Sedlacek
2000–05	Secondary Statistical Modeling With the National Assessment of Adult Literacy: Implications for the Design of the Background Questionnaire	Sheida White
2000–06	Using Telephone and Mail Surveys as a Supplement or Alternative to Door-to-Door Surveys in the Assessment of Adult Literacy	Sheida White
2000–07	"How Much Literacy is Enough?" Issues in Defining and Reporting Performance Standards for the National Assessment of Adult Literacy	Sheida White
2000–08	Evaluation of the 1992 NALS Background Survey Questionnaire: An Analysis of Uses with Recommendations for Revisions	Sheida White
2000–09	Demographic Changes and Literacy Development in a Decade	Sheida White
2001–08	Assessing the Lexile Framework: Results of a Panel Meeting	Sheida White
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom

No.	Title	NCES contact
National Assessment of Educational Progress (NAEP)		
95–12	Rural Education Data User's Guide	Samuel Peng
97–29	Can State Assessment Data be Used to Reduce State NAEP Sample Sizes?	Steven Gorman
97–30	ACT's NAEP Redesign Project: Assessment Design is the Key to Useful and Stable Assessment Results	Steven Gorman
97–31	NAEP Reconfigured: An Integrated Redesign of the National Assessment of Educational Progress	Steven Gorman
97–32	Innovative Solutions to Intractable Large Scale Assessment (Problem 2: Background Questionnaires)	Steven Gorman
97–37	Optimal Rating Procedures and Methodology for NAEP Open-ended Items	Steven Gorman
97–44	Development of a SASS 1993–94 School-Level Student Achievement Subfile: Using State Assessments and State NAEP, Feasibility Study	Michael Ross
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
1999–05	Procedures Guide for Transcript Studies	Dawn Nelson
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson
2001–07	A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)	Arnold Goldstein
2001–08	Assessing the Lexile Framework: Results of a Panel Meeting	Sheida White
2001–11	Impact of Selected Background Variables on Students' NAEP Math Performance	Arnold Goldstein
2001–13	The Effects of Accommodations on the Assessment of LEP Students in NAEP	Arnold Goldstein
2001–19	The Measurement of Home Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Graders to Questionnaire Items and Parental Assessment of the Invasiveness of These Items	Arnold Goldstein
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
2002-06	The Measurement of Instructional Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Grade Students and Teachers to Questionnaire Items	Arnold Goldstein
2003-06	NAEP Validity Studies: The Validity of Oral Accommodation in Testing	Patricia Dabbs
2003-07	NAEP Validity Studies: An Agenda for NAEP Validity Research	Patricia Dabbs
2003-08	NAEP Validity Studies: Improving the Information Value of Performance Items in Large Scale Assessments	Patricia Dabbs
2003-09	NAEP Validity Studies: Optimizing State NAEP: Issues and Possible Improvements	Patricia Dabbs
2003-10	A Content Comparison of the NAEP and PIRLS Fourth-Grade Reading Assessments	Marilyn Binkley
2003-11	NAEP Validity Studies: Reporting the Results of the National Assessment of Educational Progress	Patricia Dabbs
2003-12	NAEP Validity Studies: An Investigation of Why Students Do Not Respond to Questions	Patricia Dabbs
2003-13	NAEP Validity Studies: A Study of Equating in NAEP	Patricia Dabbs
2003-14	NAEP Validity Studies: Feasibility Studies of Two-Stage Testing in Large-Scale Educational Assessment: Implications for NAEP	Patricia Dabbs
2003-15	NAEP Validity Studies: Computer Use and Its Relation to Academic Achievement in Mathematics, Reading, and Writing	Patricia Dabbs
2003-16	NAEP Validity Studies: Implications of Electronic Technology for the NAEP Assessment	Patricia Dabbs
2003-17	NAEP Validity Studies: The Effects of Finite Sampling on State Assessment Sample Requirements	Patricia Dabbs
2003-19	NAEP Quality Assurance Checks of the 2002 Reading Assessment Results of Delaware	Janis Brown
National Education Longitudinal Study of 1988 (NELS:88)		
95–04	National Education Longitudinal Study of 1988: Second Follow-up Questionnaire Content Areas and Research Issues	Jeffrey Owings
95–05	National Education Longitudinal Study of 1988: Conducting Trend Analyses of NLS-72, HS&B, and NELS:88 Seniors	Jeffrey Owings
95–06	National Education Longitudinal Study of 1988: Conducting Cross-Cohort Comparisons Using HS&B, NAEP, and NELS:88 Academic Transcript Data	Jeffrey Owings
95–07	National Education Longitudinal Study of 1988: Conducting Trend Analyses HS&B and NELS:88 Sophomore Cohort Dropouts	Jeffrey Owings
95–12	Rural Education Data User's Guide	Samuel Peng
95–14	Empirical Evaluation of Social, Psychological, & Educational Construct Variables Used in NCES Surveys	Samuel Peng
96–03	National Education Longitudinal Study of 1988 (NELS:88) Research Framework and Issues	Jeffrey Owings

No.	Title	NCES contact
98–06	National Education Longitudinal Study of 1988 (NELS:88) Base Year through Second Follow-Up: Final Methodology Report	Ralph Lee
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
1999–05	Procedures Guide for Transcript Studies	Dawn Nelson
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson
1999–15	Projected Postsecondary Outcomes of 1992 High School Graduates	Aurora D'Amico
2001–16	Imputation of Test Scores in the National Education Longitudinal Study of 1988	Ralph Lee
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
2003–01	Mathematics, Foreign Language, and Science Coursetaking and the NELS:88 Transcript Data	Jeffrey Owings
2003–02	English Coursetaking and the NELS:88 Transcript Data	Jeffrey Owings
2003-18	Report for Computation of Balanced Repeated Replicate (BRR) Weights for the Third (NELS88:1994) and Fourth (NELS88:2000) Follow-up Surveys	Dennis Carroll
National Household Education Survey (NHES)		
95–12	Rural Education Data User's Guide	Samuel Peng
96–13	Estimation of Response Bias in the NHES:95 Adult Education Survey	Steven Kaufman
96–14	The 1995 National Household Education Survey: Reinterview Results for the Adult Education Component	Steven Kaufman
96–20	1991 National Household Education Survey (NHES:91) Questionnaires: Screener, Early Childhood Education, and Adult Education	Kathryn Chandler
96–21	1993 National Household Education Survey (NHES:93) Questionnaires: Screener, School Readiness, and School Safety and Discipline	Kathryn Chandler
96–22	1995 National Household Education Survey (NHES:95) Questionnaires: Screener, Early Childhood Program Participation, and Adult Education	Kathryn Chandler
96–29	Undercoverage Bias in Estimates of Characteristics of Adults and 0- to 2-Year-Olds in the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
96–30	Comparison of Estimates from the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
97–02	Telephone Coverage Bias and Recorded Interviews in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97–03	1991 and 1995 National Household Education Survey Questionnaires: NHES:91 Screener, NHES:91 Adult Education, NHES:95 Basic Screener, and NHES:95 Adult Education	Kathryn Chandler
97–04	Design, Data Collection, Monitoring, Interview Administration Time, and Data Editing in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97–05	Unit and Item Response, Weighting, and Imputation Procedures in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97–06	Unit and Item Response, Weighting, and Imputation Procedures in the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
97–08	Design, Data Collection, Interview Timing, and Data Editing in the 1995 National Household Education Survey	Kathryn Chandler
97–19	National Household Education Survey of 1995: Adult Education Course Coding Manual	Peter Stowe
97–20	National Household Education Survey of 1995: Adult Education Course Code Merge Files User's Guide	Peter Stowe
97–25	1996 National Household Education Survey (NHES:96) Questionnaires: Screener/Household and Library, Parent and Family Involvement in Education and Civic Involvement, Youth Civic Involvement, and Adult Civic Involvement	Kathryn Chandler
97–28	Comparison of Estimates in the 1996 National Household Education Survey	Kathryn Chandler
97–34	Comparison of Estimates from the 1993 National Household Education Survey	Kathryn Chandler
97–35	Design, Data Collection, Interview Administration Time, and Data Editing in the 1996 National Household Education Survey	Kathryn Chandler
97–38	Reinterview Results for the Parent and Youth Components of the 1996 National Household Education Survey	Kathryn Chandler
97–39	Undercoverage Bias in Estimates of Characteristics of Households and Adults in the 1996 National Household Education Survey	Kathryn Chandler
97–40	Unit and Item Response Rates, Weighting, and Imputation Procedures in the 1996 National Household Education Survey	Kathryn Chandler
98–03	Adult Education in the 1990s: A Report on the 1991 National Household Education Survey	Peter Stowe

No.	Title	NCES contact
98–10	Adult Education Participation Decisions and Barriers: Review of Conceptual Frameworks and Empirical Studies	Peter Stowe
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
National Longitudinal Study of the High School Class of 1972 (NLS-72)		
95–12	Rural Education Data User's Guide	Samuel Peng
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
National Postsecondary Student Aid Study (NPSAS)		
96–17	National Postsecondary Student Aid Study: 1996 Field Test Methodology Report	Andrew G. Malizio
2000–17	National Postsecondary Student Aid Study:2000 Field Test Methodology Report	Andrew G. Malizio
2002–03	National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), CATI Nonresponse Bias Analysis Report.	Andrew Malizio
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
2003–20	Imputation Methodology for the National Postsecondary Student Aid Study: 2004	James Griffith
National Study of Postsecondary Faculty (NSOPF)		
97–26	Strategies for Improving Accuracy of Postsecondary Faculty Lists	Linda Zimbler
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
2000–01	1999 National Study of Postsecondary Faculty (NSOPF:99) Field Test Report	Linda Zimbler
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
2002–08	A Profile of Part-time Faculty: Fall 1998	Linda Zimbler
Postsecondary Education Descriptive Analysis Reports (PEDAR)		
2000–11	Financial Aid Profile of Graduate Students in Science and Engineering	Aurora D'Amico
Private School Universe Survey (PSS)		
95–16	Intersurvey Consistency in NCES Private School Surveys	Steven Kaufman
95–17	Estimates of Expenditures for Private K–12 Schools	Stephen Broughman
96–16	Strategies for Collecting Finance Data from Private Schools	Stephen Broughman
96–26	Improving the Coverage of Private Elementary-Secondary Schools	Steven Kaufman
96–27	Intersurvey Consistency in NCES Private School Surveys for 1993–94	Steven Kaufman
97–07	The Determinants of Per-Pupil Expenditures in Private Elementary and Secondary Schools: An Exploratory Analysis	Stephen Broughman
97–22	Collection of Private School Finance Data: Development of a Questionnaire	Stephen Broughman
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
2000–15	Feasibility Report: School-Level Finance Pretest, Private School Questionnaire	Stephen Broughman
Progress in International Reading Literacy Study (PIRLS)		
2003–05	PIRLS-IEA Reading Literacy Framework: Comparative Analysis of the 1991 IEA Reading Study and the Progress in International Reading Literacy Study	Laurence Ogle
2003-10	A Content Comparison of the NAEP and PIRLS Fourth-Grade Reading Assessments	Marilyn Binkley
Recent College Graduates (RCG)		
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
Schools and Staffing Survey (SASS)		
94–01	Schools and Staffing Survey (SASS) Papers Presented at Meetings of the American Statistical Association	Dan Kasprzyk
94–02	Generalized Variance Estimate for Schools and Staffing Survey (SASS)	Dan Kasprzyk
94–03	1991 Schools and Staffing Survey (SASS) Reinterview Response Variance Report	Dan Kasprzyk
94–04	The Accuracy of Teachers' Self-reports on their Postsecondary Education: Teacher Transcript Study, Schools and Staffing Survey	Dan Kasprzyk
94–06	Six Papers on Teachers from the 1990–91 Schools and Staffing Survey and Other Related Surveys	Dan Kasprzyk
95–01	Schools and Staffing Survey: 1994 Papers Presented at the 1994 Meeting of the American Statistical Association	Dan Kasprzyk
95–02	QED Estimates of the 1990–91 Schools and Staffing Survey: Deriving and Comparing QED School Estimates with CCD Estimates	Dan Kasprzyk
95–03	Schools and Staffing Survey: 1990–91 SASS Cross-Questionnaire Analysis	Dan Kasprzyk

No.	Title	NCES contact
95–08	CCD Adjustment to the 1990–91 SASS: A Comparison of Estimates	Dan Kasprzyk
95–09	The Results of the 1993 Teacher List Validation Study (TLVS)	Dan Kasprzyk
95–10	The Results of the 1991–92 Teacher Follow-up Survey (TFS) Reinterview and Extensive Reconciliation	Dan Kasprzyk
95–11	Measuring Instruction, Curriculum Content, and Instructional Resources: The Status of Recent Work	Sharon Bobbitt & John Ralph
95–12	Rural Education Data User's Guide	Samuel Peng
95–14	Empirical Evaluation of Social, Psychological, & Educational Construct Variables Used in NCES Surveys	Samuel Peng
95–15	Classroom Instructional Processes: A Review of Existing Measurement Approaches and Their Applicability for the Teacher Follow-up Survey	Sharon Bobbitt
95–16	Intersurvey Consistency in NCES Private School Surveys	Steven Kaufman
95–18	An Agenda for Research on Teachers and Schools: Revisiting NCES' Schools and Staffing Survey	Dan Kasprzyk
96–01	Methodological Issues in the Study of Teachers' Careers: Critical Features of a Truly Longitudinal Study	Dan Kasprzyk
96–02	Schools and Staffing Survey (SASS): 1995 Selected papers presented at the 1995 Meeting of the American Statistical Association	Dan Kasprzyk
96–05	Cognitive Research on the Teacher Listing Form for the Schools and Staffing Survey	Dan Kasprzyk
96–06	The Schools and Staffing Survey (SASS) for 1998–99: Design Recommendations to Inform Broad Education Policy	Dan Kasprzyk
96–07	Should SASS Measure Instructional Processes and Teacher Effectiveness?	Dan Kasprzyk
96–09	Making Data Relevant for Policy Discussions: Redesigning the School Administrator Questionnaire for the 1998–99 SASS	Dan Kasprzyk
96–10	1998–99 Schools and Staffing Survey: Issues Related to Survey Depth	Dan Kasprzyk
96–11	Towards an Organizational Database on America's Schools: A Proposal for the Future of SASS, with comments on School Reform, Governance, and Finance	Dan Kasprzyk
96–12	Predictors of Retention, Transfer, and Attrition of Special and General Education Teachers: Data from the 1989 Teacher Followup Survey	Dan Kasprzyk
96–15	Nested Structures: District-Level Data in the Schools and Staffing Survey	Dan Kasprzyk
96–23	Linking Student Data to SASS: Why, When, How	Dan Kasprzyk
96–24	National Assessments of Teacher Quality	Dan Kasprzyk
96–25	Measures of Inservice Professional Development: Suggested Items for the 1998–1999 Schools and Staffing Survey	Dan Kasprzyk
96–28	Student Learning, Teaching Quality, and Professional Development: Theoretical Linkages, Current Measurement, and Recommendations for Future Data Collection	Mary Rollefson
97–01	Selected Papers on Education Surveys: Papers Presented at the 1996 Meeting of the American Statistical Association	Dan Kasprzyk
97–07	The Determinants of Per-Pupil Expenditures in Private Elementary and Secondary Schools: An Exploratory Analysis	Stephen Broughman
97–09	Status of Data on Crime and Violence in Schools: Final Report	Lee Hoffman
97–10	Report of Cognitive Research on the Public and Private School Teacher Questionnaires for the Schools and Staffing Survey 1993–94 School Year	Dan Kasprzyk
97–11	International Comparisons of Inservice Professional Development	Dan Kasprzyk
97–12	Measuring School Reform: Recommendations for Future SASS Data Collection	Mary Rollefson
97–14	Optimal Choice of Periodicities for the Schools and Staffing Survey: Modeling and Analysis	Steven Kaufman
97–18	Improving the Mail Return Rates of SASS Surveys: A Review of the Literature	Steven Kaufman
97–22	Collection of Private School Finance Data: Development of a Questionnaire	Stephen Broughman
97–23	Further Cognitive Research on the Schools and Staffing Survey (SASS) Teacher Listing Form	Dan Kasprzyk
97–41	Selected Papers on the Schools and Staffing Survey: Papers Presented at the 1997 Meeting of the American Statistical Association	Steve Kaufman
97–42	Improving the Measurement of Staffing Resources at the School Level: The Development of Recommendations for NCES for the Schools and Staffing Survey (SASS)	Mary Rollefson
97–44	Development of a SASS 1993–94 School-Level Student Achievement Subfile: Using State Assessments and State NAEP, Feasibility Study	Michael Ross
98–01	Collection of Public School Expenditure Data: Development of a Questionnaire	Stephen Broughman
98–02	Response Variance in the 1993–94 Schools and Staffing Survey: A Reinterview Report	Steven Kaufman
98–04	Geographic Variations in Public Schools' Costs	William J. Fowler, Jr.
98–05	SASS Documentation: 1993–94 SASS Student Sampling Problems; Solutions for Determining the Numerators for the SASS Private School (3B) Second-Stage Factors	Steven Kaufman
98–08	The Redesign of the Schools and Staffing Survey for 1999–2000: A Position Paper	Dan Kasprzyk

No.	Title	NCES contact
98–12	A Bootstrap Variance Estimator for Systematic PPS Sampling	Steven Kaufman
98–13	Response Variance in the 1994–95 Teacher Follow-up Survey	Steven Kaufman
98–14	Variance Estimation of Imputed Survey Data	Steven Kaufman
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
98–16	A Feasibility Study of Longitudinal Design for Schools and Staffing Survey	Stephen Broughman
1999–02	Tracking Secondary Use of the Schools and Staffing Survey Data: Preliminary Results	Dan Kasprzyk
1999–04	Measuring Teacher Qualifications	Dan Kasprzyk
1999–07	Collection of Resource and Expenditure Data on the Schools and Staffing Survey	Stephen Broughman
1999–08	Measuring Classroom Instructional Processes: Using Survey and Case Study Fieldtest Results to Improve Item Construction	Dan Kasprzyk
1999–10	What Users Say About Schools and Staffing Survey Publications	Dan Kasprzyk
1999–12	1993–94 Schools and Staffing Survey: Data File User's Manual, Volume III: Public-Use Codebook	Kerry Gruber
1999–13	1993–94 Schools and Staffing Survey: Data File User's Manual, Volume IV: Bureau of Indian Affairs (BIA) Restricted-Use Codebook	Kerry Gruber
1999–14	1994–95 Teacher Followup Survey: Data File User's Manual, Restricted-Use Codebook	Kerry Gruber
1999–17	Secondary Use of the Schools and Staffing Survey Data	Susan Wiley
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
2000–10	A Research Agenda for the 1999–2000 Schools and Staffing Survey	Dan Kasprzyk
2000–13	Non-professional Staff in the Schools and Staffing Survey (SASS) and Common Core of Data (CCD)	Kerry Gruber
2000–18	Feasibility Report: School-Level Finance Pretest, Public School District Questionnaire	Stephen Broughman
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
Third International Mathematics and Science Study (TIMSS)		
2001–01	Cross-National Variation in Educational Preparation for Adulthood: From Early Adolescence to Young Adulthood	Elvira Hausken
2001–05	Using TIMSS to Analyze Correlates of Performance Variation in Mathematics	Patrick Gonzales
2001–07	A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)	Arnold Goldstein
2002–01	Legal and Ethical Issues in the Use of Video in Education Research	Patrick Gonzales

This page intentionally left blank.

Listing of NCES Working Papers by Subject

No.	Title	NCES contact
Achievement (student) - mathematics		
2001–05	Using TIMSS to Analyze Correlates of Performance Variation in Mathematics	Patrick Gonzales
Adult education		
96–14	The 1995 National Household Education Survey: Reinterview Results for the Adult Education Component	Steven Kaufman
96–20	1991 National Household Education Survey (NHES:91) Questionnaires: Screener, Early Childhood Education, and Adult Education	Kathryn Chandler
96–22	1995 National Household Education Survey (NHES:95) Questionnaires: Screener, Early Childhood Program Participation, and Adult Education	Kathryn Chandler
98–03	Adult Education in the 1990s: A Report on the 1991 National Household Education Survey	Peter Stowe
98–10	Adult Education Participation Decisions and Barriers: Review of Conceptual Frameworks and Empirical Studies	Peter Stowe
1999–11	Data Sources on Lifelong Learning Available from the National Center for Education Statistics	Lisa Hudson
2000–16a	Lifelong Learning NCES Task Force: Final Report Volume I	Lisa Hudson
2000–16b	Lifelong Learning NCES Task Force: Final Report Volume II	Lisa Hudson
Adult literacy—see Literacy of adults		
American Indian – education		
1999–13	1993–94 Schools and Staffing Survey: Data File User's Manual, Volume IV: Bureau of Indian Affairs (BIA) Restricted-Use Codebook	Kerry Gruber
Assessment/achievement		
95–12	Rural Education Data User's Guide	Samuel Peng
95–13	Assessing Students with Disabilities and Limited English Proficiency	James Houser
97–29	Can State Assessment Data be Used to Reduce State NAEP Sample Sizes?	Larry Ogle
97–30	ACT's NAEP Redesign Project: Assessment Design is the Key to Useful and Stable Assessment Results	Larry Ogle
97–31	NAEP Reconfigured: An Integrated Redesign of the National Assessment of Educational Progress	Larry Ogle
97–32	Innovative Solutions to Intractable Large Scale Assessment (Problem 2: Background Questions)	Larry Ogle
97–37	Optimal Rating Procedures and Methodology for NAEP Open-ended Items	Larry Ogle
97–44	Development of a SASS 1993–94 School-Level Student Achievement Subfile: Using State Assessments and State NAEP, Feasibility Study	Michael Ross
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
2001–07	A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)	Arnold Goldstein
2001–11	Impact of Selected Background Variables on Students' NAEP Math Performance	Arnold Goldstein
2001–13	The Effects of Accommodations on the Assessment of LEP Students in NAEP	Arnold Goldstein
2001–19	The Measurement of Home Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Graders to Questionnaire Items and Parental Assessment of the Invasiveness of These Items	Arnold Goldstein
2002-05	Early Childhood Longitudinal Study-Kindergarten Class of 1998–99 (ECLS–K), Psychometric Report for Kindergarten Through First Grade	Elvira Hausken
2002-06	The Measurement of Instructional Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Grade Students and Teachers to Questionnaire Items	Arnold Goldstein
2003-19	NAEP Quality Assurance Checks of the 2002 Reading Assessment Results of Delaware	Janis Brown

No.	Title	NCES contact
Beginning students in postsecondary education		
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D'Amico
2001–04	Beginning Postsecondary Students Longitudinal Study: 1996–2001 (BPS:1996/2001) Field Test Methodology Report	Paula Knepper
Civic participation		
97–25	1996 National Household Education Survey (NHES:96) Questionnaires: Screener/Household and Library, Parent and Family Involvement in Education and Civic Involvement, Youth Civic Involvement, and Adult Civic Involvement	Kathryn Chandler
Climate of schools		
95–14	Empirical Evaluation of Social, Psychological, & Educational Construct Variables Used in NCES Surveys	Samuel Peng
Cost of education indices		
94–05	Cost-of-Education Differentials Across the States	William J. Fowler, Jr.
Course-taking		
95–12	Rural Education Data User's Guide	Samuel Peng
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
1999–05	Procedures Guide for Transcript Studies	Dawn Nelson
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson
2003–01	Mathematics, Foreign Language, and Science Coursetaking and the NELS:88 Transcript Data	Jeffrey Owings
2003–02	English Coursetaking and the NELS:88 Transcript Data	Jeffrey Owings
Crime		
97–09	Status of Data on Crime and Violence in Schools: Final Report	Lee Hoffman
Curriculum		
95–11	Measuring Instruction, Curriculum Content, and Instructional Resources: The Status of Recent Work	Sharon Bobbitt & John Ralph
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
Customer service		
1999–10	What Users Say About Schools and Staffing Survey Publications	Dan Kasprzyk
2000–02	Coordinating NCES Surveys: Options, Issues, Challenges, and Next Steps	Valena Plisko
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
Data quality		
97–13	Improving Data Quality in NCES: Database-to-Report Process	Susan Ahmed
2001–11	Impact of Selected Background Variables on Students' NAEP Math Performance	Arnold Goldstein
2001–13	The Effects of Accommodations on the Assessment of LEP Students in NAEP	Arnold Goldstein
2001–19	The Measurement of Home Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Graders to Questionnaire Items and Parental Assessment of the Invasiveness of These Items	Arnold Goldstein
2002-06	The Measurement of Instructional Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Grade Students and Teachers to Questionnaire Items	Arnold Goldstein
2003-19	NAEP Quality Assurance Checks of the 2002 Reading Assessment Results of Delaware	Janis Brown
Data warehouse		
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk

No.	Title	NCES contact
Design effects		
2000–03	Strengths and Limitations of Using SUDAAN, Stata, and WesVarPC for Computing Variances from NCES Data Sets	Ralph Lee
Dropout rates, high school		
95–07	National Education Longitudinal Study of 1988: Conducting Trend Analyses HS&B and NELS:88 Sophomore Cohort Dropouts	Jeffrey Owings
Early childhood education		
96–20	1991 National Household Education Survey (NHES:91) Questionnaires: Screener, Early Childhood Education, and Adult Education	Kathryn Chandler
96–22	1995 National Household Education Survey (NHES:95) Questionnaires: Screener, Early Childhood Program Participation, and Adult Education	Kathryn Chandler
97–24	Formulating a Design for the ECLS: A Review of Longitudinal Studies	Jerry West
97–36	Measuring the Quality of Program Environments in Head Start and Other Early Childhood Programs: A Review and Recommendations for Future Research	Jerry West
1999–01	A Birth Cohort Study: Conceptual and Design Considerations and Rationale	Jerry West
2001–02	Measuring Father Involvement in Young Children's Lives: Recommendations for a Fatherhood Module for the ECLS-B	Jerry West
2001–03	Measures of Socio-Emotional Development in Middle School	Elvira Hausken
2001–06	Papers from the Early Childhood Longitudinal Studies Program: Presented at the 2001 AERA and SRCD Meetings	Jerry West
2002-05	Early Childhood Longitudinal Study-Kindergarten Class of 1998–99 (ECLS–K), Psychometric Report for Kindergarten Through First Grade	Elvira Hausken
Educational attainment		
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D'Amico
2001–15	Baccalaureate and Beyond Longitudinal Study: 2000/01 Follow-Up Field Test Methodology Report	Andrew G. Malizio
Educational research		
2000–02	Coordinating NCES Surveys: Options, Issues, Challenges, and Next Steps	Valena Plisko
2002–01	Legal and Ethical Issues in the Use of Video in Education Research	Patrick Gonzales
Eighth-graders		
2001–05	Using TIMSS to Analyze Correlates of Performance Variation in Mathematics	Patrick Gonzales
Employment		
96–03	National Education Longitudinal Study of 1988 (NELS:88) Research Framework and Issues	Jeffrey Owings
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D'Amico
2000–16a	Lifelong Learning NCES Task Force: Final Report Volume I	Lisa Hudson
2000–16b	Lifelong Learning NCES Task Force: Final Report Volume II	Lisa Hudson
2001–01	Cross-National Variation in Educational Preparation for Adulthood: From Early Adolescence to Young Adulthood	Elvira Hausken
Employment – after college		
2001–15	Baccalaureate and Beyond Longitudinal Study: 2000/01 Follow-Up Field Test Methodology Report	Andrew G. Malizio
Engineering		
2000–11	Financial Aid Profile of Graduate Students in Science and Engineering	Aurora D'Amico
Enrollment – after college		
2001–15	Baccalaureate and Beyond Longitudinal Study: 2000/01 Follow-Up Field Test Methodology Report	Andrew G. Malizio
Faculty – higher education		
97–26	Strategies for Improving Accuracy of Postsecondary Faculty Lists	Linda Zimbler
2000–01	1999 National Study of Postsecondary Faculty (NSOPF:99) Field Test Report	Linda Zimbler
2002–08	A Profile of Part-time Faculty: Fall 1998	Linda Zimbler

No.	Title	NCES contact
Fathers – role in education		
2001–02	Measuring Father Involvement in Young Children's Lives: Recommendations for a Fatherhood Module for the ECLS-B	Jerry West
Finance – elementary and secondary schools		
94–05	Cost-of-Education Differentials Across the States	William J. Fowler, Jr.
96–19	Assessment and Analysis of School-Level Expenditures	William J. Fowler, Jr.
98–01	Collection of Public School Expenditure Data: Development of a Questionnaire	Stephen Broughman
1999–07	Collection of Resource and Expenditure Data on the Schools and Staffing Survey	Stephen Broughman
1999–16	Measuring Resources in Education: From Accounting to the Resource Cost Model Approach	William J. Fowler, Jr.
2000–18	Feasibility Report: School-Level Finance Pretest, Public School District Questionnaire	Stephen Broughman
Finance – postsecondary		
97–27	Pilot Test of IPEDS Finance Survey	Peter Stowe
2000–14	IPEDS Finance Data Comparisons Under the 1997 Financial Accounting Standards for Private, Not-for-Profit Institutes: A Concept Paper	Peter Stowe
Finance – private schools		
95–17	Estimates of Expenditures for Private K–12 Schools	Stephen Broughman
96–16	Strategies for Collecting Finance Data from Private Schools	Stephen Broughman
97–07	The Determinants of Per-Pupil Expenditures in Private Elementary and Secondary Schools: An Exploratory Analysis	Stephen Broughman
97–22	Collection of Private School Finance Data: Development of a Questionnaire	Stephen Broughman
1999–07	Collection of Resource and Expenditure Data on the Schools and Staffing Survey	Stephen Broughman
2000–15	Feasibility Report: School-Level Finance Pretest, Private School Questionnaire	Stephen Broughman
Geography		
98–04	Geographic Variations in Public Schools' Costs	William J. Fowler, Jr.
Graduate students		
2000–11	Financial Aid Profile of Graduate Students in Science and Engineering	Aurora D'Amico
Graduates of postsecondary education		
2001–15	Baccalaureate and Beyond Longitudinal Study: 2000/01 Follow-Up Field Test Methodology Report	Andrew G. Malizio
Imputation		
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meeting	Dan Kasprzyk
2001–10	Comparison of Proc Impute and Schafer's Multiple Imputation Software	Sam Peng
2001–16	Imputation of Test Scores in the National Education Longitudinal Study of 1988	Ralph Lee
2001–17	A Study of Imputation Algorithms	Ralph Lee
2001–18	A Study of Variance Estimation Methods	Ralph Lee
2003–20	Imputation Methodology for the National Postsecondary Student Aid Study: 2004	James Griffith
Inflation		
97–43	Measuring Inflation in Public School Costs	William J. Fowler, Jr.
Institution data		
2000–01	1999 National Study of Postsecondary Faculty (NSOPF:99) Field Test Report	Linda Zimbler
Instructional resources and practices		
95–11	Measuring Instruction, Curriculum Content, and Instructional Resources: The Status of Recent Work	Sharon Bobbitt & John Ralph
1999–08	Measuring Classroom Instructional Processes: Using Survey and Case Study Field Test Results to Improve Item Construction	Dan Kasprzyk
International comparisons		
97–11	International Comparisons of Inservice Professional Development	Dan Kasprzyk
97–16	International Education Expenditure Comparability Study: Final Report, Volume I	Shelley Burns

No.	Title	NCES contact
97–17	International Education Expenditure Comparability Study: Final Report, Volume II, Quantitative Analysis of Expenditure Comparability	Shelley Burns
2001–01	Cross-National Variation in Educational Preparation for Adulthood: From Early Adolescence to Young Adulthood	Elvira Hausken
2001–07	A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)	Arnold Goldstein
International comparisons – math and science achievement		
2001–05	Using TIMSS to Analyze Correlates of Performance Variation in Mathematics	Patrick Gonzales
Libraries		
94–07	Data Comparability and Public Policy: New Interest in Public Library Data Papers Presented at Meetings of the American Statistical Association	Carrol Kindel
97–25	1996 National Household Education Survey (NHES:96) Questionnaires: Screener/Household and Library, Parent and Family Involvement in Education and Civic Involvement, Youth Civic Involvement, and Adult Civic Involvement	Kathryn Chandler
Limited English Proficiency		
95–13	Assessing Students with Disabilities and Limited English Proficiency	James Houser
2001–11	Impact of Selected Background Variables on Students' NAEP Math Performance	Arnold Goldstein
2001–13	The Effects of Accommodations on the Assessment of LEP Students in NAEP	Arnold Goldstein
Literacy of adults		
98–17	Developing the National Assessment of Adult Literacy: Recommendations from Stakeholders	Sheida White
1999–09a	1992 National Adult Literacy Survey: An Overview	Alex Sedlacek
1999–09b	1992 National Adult Literacy Survey: Sample Design	Alex Sedlacek
1999–09c	1992 National Adult Literacy Survey: Weighting and Population Estimates	Alex Sedlacek
1999–09d	1992 National Adult Literacy Survey: Development of the Survey Instruments	Alex Sedlacek
1999–09e	1992 National Adult Literacy Survey: Scaling and Proficiency Estimates	Alex Sedlacek
1999–09f	1992 National Adult Literacy Survey: Interpreting the Adult Literacy Scales and Literacy Levels	Alex Sedlacek
1999–09g	1992 National Adult Literacy Survey: Literacy Levels and the Response Probability Convention	Alex Sedlacek
1999–11	Data Sources on Lifelong Learning Available from the National Center for Education Statistics	Lisa Hudson
2000–05	Secondary Statistical Modeling With the National Assessment of Adult Literacy: Implications for the Design of the Background Questionnaire	Sheida White
2000–06	Using Telephone and Mail Surveys as a Supplement or Alternative to Door-to-Door Surveys in the Assessment of Adult Literacy	Sheida White
2000–07	"How Much Literacy is Enough?" Issues in Defining and Reporting Performance Standards for the National Assessment of Adult Literacy	Sheida White
2000–08	Evaluation of the 1992 NALS Background Survey Questionnaire: An Analysis of Uses with Recommendations for Revisions	Sheida White
2000–09	Demographic Changes and Literacy Development in a Decade	Sheida White
2001–08	Assessing the Lexile Framework: Results of a Panel Meeting	Sheida White
Literacy of adults – international		
97–33	Adult Literacy: An International Perspective	Marilyn Binkley
Mathematics		
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
1999–08	Measuring Classroom Instructional Processes: Using Survey and Case Study Field Test Results to Improve Item Construction	Dan Kasprzyk
2001–05	Using TIMSS to Analyze Correlates of Performance Variation in Mathematics	Patrick Gonzales
2001–07	A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)	Arnold Goldstein
2001–11	Impact of Selected Background Variables on Students' NAEP Math Performance	Arnold Goldstein

No.	Title	NCES contact
2002-06	The Measurement of Instructional Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Grade Students and Teachers to Questionnaire Items	
Parental involvement in education		
96–03	National Education Longitudinal Study of 1988 (NELS:88) Research Framework and Issues	Jeffrey Owings
97–25	1996 National Household Education Survey (NHES:96) Questionnaires: Screener/Household and Library, Parent and Family Involvement in Education and Civic Involvement, Youth Civic Involvement, and Adult Civic Involvement	Kathryn Chandler
1999–01	A Birth Cohort Study: Conceptual and Design Considerations and Rationale	Jerry West
2001–06	Papers from the Early Childhood Longitudinal Studies Program: Presented at the 2001 AERA and SRCD Meetings	Jerry West
2001–19	The Measurement of Home Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Graders to Questionnaire Items and Parental Assessment of the Invasiveness of These Items	Arnold Goldstein
Participation rates		
98–10	Adult Education Participation Decisions and Barriers: Review of Conceptual Frameworks and Empirical Studies	Peter Stowe
Postsecondary education		
1999–11	Data Sources on Lifelong Learning Available from the National Center for Education Statistics	Lisa Hudson
2000–16a	Lifelong Learning NCES Task Force: Final Report Volume I	Lisa Hudson
2000–16b	Lifelong Learning NCES Task Force: Final Report Volume II	Lisa Hudson
2003–20	Imputation Methodology for the National Postsecondary Student Aid Study: 2004	James Griffith
Postsecondary education – persistence and attainment		
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D'Amico
1999–15	Projected Postsecondary Outcomes of 1992 High School Graduates	Aurora D'Amico
Postsecondary education – staff		
97–26	Strategies for Improving Accuracy of Postsecondary Faculty Lists	Linda Zimbler
2000–01	1999 National Study of Postsecondary Faculty (NSOPF:99) Field Test Report	Linda Zimbler
2002–08	A Profile of Part-time Faculty: Fall 1998	Linda Zimbler
Principals		
2000–10	A Research Agenda for the 1999–2000 Schools and Staffing Survey	Dan Kasprzyk
Private schools		
96–16	Strategies for Collecting Finance Data from Private Schools	Stephen Broughman
97–07	The Determinants of Per-Pupil Expenditures in Private Elementary and Secondary Schools: An Exploratory Analysis	Stephen Broughman
97–22	Collection of Private School Finance Data: Development of a Questionnaire	Stephen Broughman
2000–13	Non-professional Staff in the Schools and Staffing Survey (SASS) and Common Core of Data (CCD)	Kerry Gruber
2000–15	Feasibility Report: School-Level Finance Pretest, Private School Questionnaire	Stephen Broughman
Projections of education statistics		
1999–15	Projected Postsecondary Outcomes of 1992 High School Graduates	Aurora D'Amico
Public school finance		
1999–16	Measuring Resources in Education: From Accounting to the Resource Cost Model Approach	William J. Fowler, Jr.
2000–18	Feasibility Report: School-Level Finance Pretest, Public School District Questionnaire	Stephen Broughman
Public schools		
97–43	Measuring Inflation in Public School Costs	William J. Fowler, Jr.
98–01	Collection of Public School Expenditure Data: Development of a Questionnaire	Stephen Broughman
98–04	Geographic Variations in Public Schools' Costs	William J. Fowler, Jr.
1999–02	Tracking Secondary Use of the Schools and Staffing Survey Data: Preliminary Results	Dan Kasprzyk

No.	Title	NCES contact
2000–12	Coverage Evaluation of the 1994–95 Public Elementary/Secondary School Universe Survey	Beth Young
2000–13	Non-professional Staff in the Schools and Staffing Survey (SASS) and Common Core of Data (CCD)	Kerry Gruber
2002–02	Locale Codes 1987 - 2000	Frank Johnson
Public schools – secondary		
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
Reform, educational		
96–03	National Education Longitudinal Study of 1988 (NELS:88) Research Framework and Issues	Jeffrey Owings
Response rates		
98–02	Response Variance in the 1993–94 Schools and Staffing Survey: A Reinterview Report	Steven Kaufman
School districts		
2000–10	A Research Agenda for the 1999–2000 Schools and Staffing Survey	Dan Kasprzyk
School districts, public		
98–07	Decennial Census School District Project Planning Report	Tai Phan
1999–03	Evaluation of the 1996–97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle	Beth Young
School districts, public – demographics of		
96–04	Census Mapping Project/School District Data Book	Tai Phan
Schools		
97–42	Improving the Measurement of Staffing Resources at the School Level: The Development of Recommendations for NCES for the Schools and Staffing Survey (SASS)	Mary Rollefson
98–08	The Redesign of the Schools and Staffing Survey for 1999–2000: A Position Paper	Dan Kasprzyk
1999–03	Evaluation of the 1996–97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle	Beth Young
2000–10	A Research Agenda for the 1999–2000 Schools and Staffing Survey	Dan Kasprzyk
2002–02	Locale Codes 1987 – 2000	Frank Johnson
Schools – safety and discipline		
97–09	Status of Data on Crime and Violence in Schools: Final Report	Lee Hoffman
Science		
2000–11	Financial Aid Profile of Graduate Students in Science and Engineering	Aurora D'Amico
2001–07	A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)	Arnold Goldstein
Software evaluation		
2000–03	Strengths and Limitations of Using SUDAAN, Stata, and WesVarPC for Computing Variances from NCES Data Sets	Ralph Lee
Staff		
97–42	Improving the Measurement of Staffing Resources at the School Level: The Development of Recommendations for NCES for the Schools and Staffing Survey (SASS)	Mary Rollefson
98–08	The Redesign of the Schools and Staffing Survey for 1999–2000: A Position Paper	Dan Kasprzyk
Staff – higher education institutions		
97–26	Strategies for Improving Accuracy of Postsecondary Faculty Lists	Linda Zimbler

No.	Title	NCES contact
2002–08	A Profile of Part-time Faculty: Fall 1998	Linda Zimbler
Staff – nonprofessional		
2000–13	Non-professional Staff in the Schools and Staffing Survey (SASS) and Common Core of Data (CCD)	Kerry Gruber
State		
1999–03	Evaluation of the 1996–97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle	Beth Young
2003-19	NAEP Quality Assurance Checks of the 2002 Reading Assessment Results of Delaware	Janis Brown
Statistical methodology		
97–21	Statistics for Policymakers or Everything You Wanted to Know About Statistics But Thought You Could Never Understand	Susan Ahmed
2003–20	Imputation Methodology for the National Postsecondary Student Aid Study: 2004	James Griffith
Statistical standards and methodology		
2001–05	Using TIMSS to Analyze Correlates of Performance Variation in Mathematics	Patrick Gonzales
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
Students with disabilities		
95–13	Assessing Students with Disabilities and Limited English Proficiency	James Houser
2001–13	The Effects of Accommodations on the Assessment of LEP Students in NAEP	Arnold Goldstein
Survey methodology		
96–17	National Postsecondary Student Aid Study: 1996 Field Test Methodology Report	Andrew G. Malizio
97–15	Customer Service Survey: Common Core of Data Coordinators	Lee Hoffman
97–35	Design, Data Collection, Interview Administration Time, and Data Editing in the 1996 National Household Education Survey	Kathryn Chandler
98–06	National Education Longitudinal Study of 1988 (NELS:88) Base Year through Second Follow-Up: Final Methodology Report	Ralph Lee
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D'Amico
98–16	A Feasibility Study of Longitudinal Design for Schools and Staffing Survey	Stephen Broughman
1999–07	Collection of Resource and Expenditure Data on the Schools and Staffing Survey	Stephen Broughman
1999–17	Secondary Use of the Schools and Staffing Survey Data	Susan Wiley
2000–01	1999 National Study of Postsecondary Faculty (NSOPF:99) Field Test Report	Linda Zimbler
2000–02	Coordinating NCES Surveys: Options, Issues, Challenges, and Next Steps	Valena Plisko
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
2000–12	Coverage Evaluation of the 1994–95 Public Elementary/Secondary School Universe Survey	Beth Young
2000–17	National Postsecondary Student Aid Study:2000 Field Test Methodology Report	Andrew G. Malizio
2001–04	Beginning Postsecondary Students Longitudinal Study: 1996–2001 (BPS:1996/2001) Field Test Methodology Report	Paula Knepper
2001–07	A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)	Arnold Goldstein
2001–11	Impact of Selected Background Variables on Students' NAEP Math Performance	Arnold Goldstein
2001–13	The Effects of Accommodations on the Assessment of LEP Students in NAEP	Arnold Goldstein
2001–19	The Measurement of Home Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Graders to Questionnaire Items and Parental Assessment of the Invasiveness of These Items	Arnold Goldstein
2002–01	Legal and Ethical Issues in the Use of Video in Education Research	Patrick Gonzales
2002–02	Locale Codes 1987 - 2000	Frank Johnson
2002–03	National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), CATI Nonresponse Bias Analysis Report.	Andrew Malizio

No.	Title	NCES contact
2002-06	The Measurement of Instructional Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Grade Students and Teachers to Questionnaire Items	Arnold Goldstein
2003-03	Education Longitudinal Study: 2002 (ELS: 2002) Field Test Report	Jeffrey Owings
Teachers		
98–13	Response Variance in the 1994–95 Teacher Follow-up Survey	Steven Kaufman
1999–14	1994–95 Teacher Followup Survey: Data File User's Manual, Restricted-Use Codebook	Kerry Gruber
2000–10	A Research Agenda for the 1999–2000 Schools and Staffing Survey	Dan Kasprzyk
Teachers – instructional practices of		
98–08	The Redesign of the Schools and Staffing Survey for 1999–2000: A Position Paper	Dan Kasprzyk
2002-06	The Measurement of Instructional Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Grade Students and Teachers to Questionnaire Items	Arnold Goldstein
Teachers – opinions regarding safety		
98–08	The Redesign of the Schools and Staffing Survey for 1999–2000: A Position Paper	Dan Kasprzyk
Teachers – performance evaluations		
1999–04	Measuring Teacher Qualifications	Dan Kasprzyk
Teachers – qualifications of		
1999–04	Measuring Teacher Qualifications	Dan Kasprzyk
Teachers – salaries of		
94–05	Cost-of-Education Differentials Across the States	William J. Fowler, Jr.
Training		
2000–16a	Lifelong Learning NCES Task Force: Final Report Volume I	Lisa Hudson
2000–16b	Lifelong Learning NCES Task Force: Final Report Volume II	Lisa Hudson
Variance estimation		
2000–03	Strengths and Limitations of Using SUDAAN, Stata, and WesVarPC for Computing Variances from NCES Data Sets	Ralph Lee
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
2001–18	A Study of Variance Estimation Methods	Ralph Lee
2003-18	Report for Computation of Balanced Repeated Replicate (BRR) Weights for the Third (NELS88:1994) and Fourth (NELS88:2000) Follow-up Surveys	Dennis Carroll
2003–20	Imputation Methodology for the National Postsecondary Student Aid Study: 2004	James Griffith
Violence		
97–09	Status of Data on Crime and Violence in Schools: Final Report	Lee Hoffman
Vocational education		
95–12	Rural Education Data User's Guide	Samuel Peng
1999–05	Procedures Guide for Transcript Studies	Dawn Nelson
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson

U.S. Department of Education
ED Pubs
8242-B Sandy Court
Jessup, MD 20794-1398