U.S. Department of Education
Institute of Education Sciences
NCES 2006-180

2004 National Postsecondary Student Aid Study (NPSAS:04) Full-Scale Methodology Report

Technical Report

2004 National Postsecondary Student Aid Study (NPSAS:04) Full-Scale Methodology Report

Technical Report

U.S. Department of Education
Institute of Education Sciences
NCES 2006-180

June 2006

Melissa Cominole
Peter Siegel
Kristin Dudley
David Roe
Theresa Gilligan
RTI International

James Griffith
Project Officer
National Center for Education Statistics

*RTI International is a trade name of Research Triangle Institute.

U.S. Department of Education
Margaret Spellings
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Education Statistics
Mark Schneider
Commissioner

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, Congress, state and 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

June 2006

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

Cominole, M., Siegel, P., Dudley, K., Roe, D., and Gilligan, T. *2004 National Postsecondary Student Aid Study (NPSAS:04) Full Scale Methodology Report* (NCES 2006–180). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

To download, view, and print the report as a PDF file, go to http://nces.ed.gov/pubsearch/pubinfo.asp?pubid=2006-180.

Contact:
Aurora D'Amico
aurora.d'amico@ed.gov
(202) 502-7334

Executive Summary

Introduction

The 2004 National Postsecondary Student Aid Study (NPSAS:04), conducted for the U.S. Department of Education's National Center for Education Statistics (NCES), collected comprehensive data regarding how students and their families pay for postsecondary education. The primary objective of NPSAS:04 is to produce reliable national estimates of characteristics related to financial aid for postsecondary students. NPSAS:04 also served as the base year of data collection for the Beginning Postsecondary Students Longitudinal Study (BPS), which will follow a cohort of students from the start of their postsecondary education and collect further data from them in 2006 and 2009.

For the first time, NPSAS:04 was conducted as the student component study of the 2004 National Study of Faculty and Students (NSoFaS:04). The faculty component—the 2004 National Study of Postsecondary Faculty (NSOPF:04)—is primarily a separate study, with the exception of institutional sampling and contacting. Historically, there has been considerable overlap in the institutions selected for participation in NPSAS and NSOPF; therefore, institutional sampling and contacting activities for both studies were coordinated in order to minimize response burden on institutions and to realize data collection efficiencies.

This report only describes the methodology and findings of NPSAS:04, which took place during the 2003–04 school year. The methodology and findings of NSOPF:04 are provided in a separate report.

Sample Design

The NPSAS:04 target population consists of all eligible students enrolled at any time between July 1, 2003 and June 30, 2004 in postsecondary institutions in the United States or Puerto Rico which had signed Title IV participation agreements with the U.S. Department of Education making them eligible for the federal student aid programs (Title IV institutions). NPSAS:04 is based on a nationally representative sample of all students (aided and nonaided) in those institutions. The institutions sampled represented all types and levels of postsecondary institutions in the United States, including public, private for-profit, and private not-for-profit institutions, at the 4-year, 2-year, and less-than-2-year levels. In the institutional sample, 1,670 institutions[1] were selected. Of these, 1,630 were determined to be eligible for NPSAS:04.[2] Enrollment lists were obtained from 1,360 of the 1,630 eligible institutions.

Approximately 109,210 undergraduate, graduate, and first-professional students enrolled in postsecondary education between July 1, 2003, and April 30, 2004, comprised the student sample, with special concern for the accurate sampling of students eligible to participate in the BPS longitudinal studies in the future. Students were selected on a flow basis from the institutions providing lists. Of the 109,210 students sampled, 8,200 were determined to be ineligible for the study, resulting in 101,010 eligible student sample members.

[1] The numbers appearing in the tables and text of this report have been rounded to the nearest tens to maintain the confidentiality of study respondents. However, percentages are based on unrounded numbers.
[2] 1,080 of these institutions were also included in the NSOPF institutional sample.

Study Respondents

Student-level data for NPSAS:04 were collected from a variety of sources, including student records (using computer-assisted data entry [CADE]), student interviews, and extant federal and private databases (CPS, and National Student Loan Data System [NSLDS]). For NPSAS:04, a definition of the minimum data requirements, regardless of source, to be considered a study respondent was adopted. About 90,750 of 101,010 eligible sample students had sufficient data across sources to be classified as study respondents, for a weighted response rate of 91 percent. Among the 90,750 study respondents, 92 percent were classified as CADE respondents and 70 percent were student interview respondents. The match rates to the other data services are also discussed.

Instrumentation

Unlike in previous NPSAS cycles, the NPSAS:04 student instrument was designed as a web-based instrument to be used both for self-administered "interviews" via the Web and by telephone interviewers. In addition, a study website was developed for access to the self-administered interview and to provide sample members with additional information about the study.

The instrument was designed to accommodate the mixed-mode data collection approach and to ensure the collection of the highest quality data. Design considerations included the following: appropriate question wording for both self-administered and telephone interviews; the provision of extensive help text to assist self-administered respondents and telephone interviewers; and pop-up boxes indicating out-of-range values.

The instrument consisted of six sections grouped by topic. The first section determined student eligibility for the NPSAS:04 study and the future BPS study, and obtained enrollment history. The second section contained questions relating to student expenses and financial aid. Included in this section were items regarding employment at the NPSAS institution, such as work-study participation, assistantships, and fellowships. Section three focused on employment and finances. Educational experiences, such as courses taken and admission test scores, were included in the fourth section, as well as educational experience items specific only to BPS respondents. The fifth section of the interview gathered background and demographic information about students and their family members. The final section, applicable only to BPS respondents, requested contacting information in order to make subsequent follow-up contact with them easier for future surveys.

Data Collection Design and Outcomes

Training

Training programs were developed for different types of project staff: institutional contactors, field data collectors for student record abstraction, help desk operators, and telephone interviewers. Institution contactors were trained to work with institutional staff to inform them of the nature of the study and to gain institutional participation. Training for field data collectors for student record abstraction emphasized the use of the various systems to monitor and transfer data. It also focused on the nature of the study and the processes associated with financial aid from an institutional perspective. Help desk operators received specific training on "frequently

asked questions" regarding the instrument and technical issues related to completion of the instrument via the Web. Help desk operators were also trained to conduct the student interview when requested by sample members. Programs on successfully locating and interviewing sample members were developed for all telephone interviewers. Topics covered in telephone interviewer training included administrative procedures required for case management; quality control of interactions with sample members, parents, and other contacts; the purpose of NPSAS:04 and the uses of the data to be collected; and the organization and operation of the web-based student instrument to be used in data collection.

Institutional Contacting

Once institutions were sampled, attempts were made to contact the chief administrator to verify institutional eligibility, solicit participation, and request the appointment of an institutional coordinator to oversee data collection within the institution. Institutional coordinators were asked to provide lists or data files of all eligible students enrolled at any time between July 1, 2003, and April 30, 2004. Several checks on quality and completeness of student lists were implemented prior to sampling students from each institution. Of the 1,630 eligible institutions sampled for the field test, about 1,360 provided lists, resulting in an overall institutional participation rate of about 80 percent (weighted).

Institutional Record Abstraction

A web-based CADE software system was used for the abstraction of student records from institutions. Institutions were given the option of completing CADE using their own staff, or, upon request, having an RTI International (RTI) field data collector complete the record abstraction process at the institution. Prior to the initialization of the CADE software system for an institution, records for all students sampled from a school were requested from the U.S. Department of Education's Central Processing System (CPS), which contains financial aid application data. This information was preloaded into the CADE system to provide edit checks for the data entered by an institution. The CADE system consisted of three sections focusing on eight topics: locating information, demographic characteristics, admissions tests, enrollment, tuition, financial aid awards, needs analysis, and institutional student information records (ISIRs). Of the 1,360 eligible and participating institutions, about 1,300 institutions provided information for about 88,920 students. The institutional and student-level weighted response rates for record abstraction were 96 percent and 92 percent, respectively.

Student Locating and Interviewing

The NPSAS:04 data collection design involved initial locating of sample members, providing an opportunity for the student to complete the self-administered interview via the Web, following up with Web nonrespondents after 4 weeks, and attempting to conduct a telephone interview with them if necessary. Upon receipt of student enrollment lists, batch-locating activities were implemented to update address and telephone activities. Sources for this task included the CPS, the U.S. Postal Service National Change of Address (NCOA) system, and Telematch. Students were then sent a notification mailing containing a lead letter, informational brochure, and username and password for completing the interview via the Web. Telephone contact began for self-administered Web nonrespondents 4 weeks after the initial mailing. Locating and tracing activities by telephone interviewers occurred simultaneously with efforts to

gain cooperation from sample members. When all tracing options were exhausted by the interviewer, cases were sent to RTI's Call Center Services (CCS) Tracing Services. Cases for which further contacting information was obtained were sent back for contact by telephone interviewers; those for whom no additional information could be obtained were finalized as unlocatable.

As discussed earlier, there were 90,750 study respondents among the 101,010 eligible sample members. About 62,220 completed the student interview, for a weighted response rate of 71 percent. Among those who completed the student interview, about 25,000 BPS respondents eligible for the longitudinal follow-up studies (BPS:04/06 and BPS:04/09) were identified.[3] Of all completed student interviews, about 53 percent (weighted) were completed with a telephone interviewer, and 47 percent (weighted) were completed via self-administration over the Web. The average overall time to complete the student interview for all respondents was about 27 minutes, regardless of the mode of completion.

Evaluation of Operations and Data Quality

Evaluations[4] of operations and procedures focused on the newly introduced joint institutional contacting endeavor, the timeline for data collection from both institutions (CADE) and students (self-administered and interviewer-administered), tracing and locating procedures, refusal conversion efforts, the effectiveness of incentives for increasing early response via the Web and for refusal conversion, and the length of the student interview. Evaluations of data quality included an examination of items with high rates of missing data, use of online help text, item-level nonresponse conversion efforts, and question delivery and data entry quality control procedures.

Analysis Weights

Cross-sectional weights were developed for analyzing respondents to the NPSAS:04 interview. Variances were computed using the Taylor series and bootstrap techniques. Weighted response rates, nonresponse bias analyses, and survey design effect tables are also provided.

Data Files

Throughout the data collection period, data were processed and examined for quality control purposes. Following completion of all study data collection, separate Data Analysis System (DAS) files were created for undergraduate and graduate/first-professional students. The first DASs, both undergraduate and graduate/first-professional, were adjudicated and approved for public release in February 2005. The primary analysis file, from which the study DASs were constructed, contains data for approximately 90,750 study respondents,[5] including about 35,510 first-time beginner (FTB) students, 44,340 other undergraduates, and 10,890 graduate and first-professional students. The primary analysis file contains over 500 variables, developed from multiple sources (including student interviews, institutional records, and extant data sources).

[3] Institutions identified all potential first time beginners (FTBs) as they prepared enrollment lists for sampling purposes. Eligibility for inclusion in the BPS:04 cohort was confirmed for those who completed the student interview. Study respondents who did not complete the student interview are retained on the data file as potential FTBs and their eligibility will be confirmed as part of the follow-up interview.

[4] All comparisons have been tested using a significance level of 0.05.

[5] Study respondents are those who met the minimum data requirements regardless of data source.

The survey data files used to create variables in the DASs, and the associated electronic codebooks and file documentation, are available to researchers who have obtained a restricted data license from NCES.

Products

NPSAS:04 reports or data products that have or will be published include the following:

- *2004 National Postsecondary Student Aid Study (NPSAS:04): Student Financial Aid Estimates for 2003–04*. This E.D. TAB is the first publication based on the 2003–04 data. The E.D. TAB describes the percentages of students receiving various types of financial aid and average amounts received, by type of institution attended, attendance pattern, dependency status, and income level.
- *2004 National Postsecondary Student Aid Study (NPSAS:04): Undergraduate Financial Aid Estimates for 2003–04 by Type of Institution*. This E.D. TAB is the second publication based on the 2003–04 data. This E.D. TAB focuses only on undergraduates, including separate tables for those who attended public 4-year, private-not-for-profit 4-year, public 2-year, or private for-profit postsecondary institutions during the 2003–04 academic year. It describes average tuition and fees, average total price of attendance, and the percentages of undergraduates receiving various types and combinations of financial aid and average amounts received, with a particular focus on grants and loans.
- *2004 National Postsecondary Student Aid Study (NPSAS:04): Undergraduate Data Analysis System*. The NPSAS:04 Undergraduate DAS contains the data on a sample of about 80,000 undergraduates who were enrolled at any time between July 1, 2003, and June 30, 2004, in about 1,400 postsecondary institutions. It represents all undergraduate students enrolled in postsecondary institutions in the 50 states, the District of Columbia, and Puerto Rico that were eligible to participate in the federal financial aid programs in Title IV of the Higher Education Act.
- *2004 National Postsecondary Student Aid Study (NPSAS:04): Graduate Data Analysis System*. The NPSAS:04 Graduate DAS contains the data on a sample of about 11,000 graduate students who were enrolled at any time between July 1, 2003, and June 30, 2004, in about 1,400 postsecondary institutions. It represents all graduate students enrolled in postsecondary institutions in the 50 states, the District of Columbia, and Puerto Rico that were eligible to participate in the federal financial aid programs in Title IV of the Higher Education Act.
- *Profile of Undergraduates in U.S. Postsecondary Education Institutions: 2003–04.* Describes the demographic and enrollment characteristics of undergraduate students.
- *Student Financing of Undergraduate Education: 2003–04.* Focuses on undergraduate tuition, total price of attendance, types and sources of financial aid received, net price, financial aid need, and unmet need.
- *Student Financing of Graduate and First-Professional Education: 2003–04*. Describes the demographic and enrollment characteristics of graduate and first-professional students and the types and sources of financial aid received.

This page left intentionally blank.

Foreword

This report describes and evaluates the methods and procedures used in the 2004 National Postsecondary Student Aid Study (NPSAS:04), the student component of the 2004 National Study of Faculty and Students (NSoFaS:04). NPSAS:04 included important changes from previous NPSAS studies. One of the most significant changes was the fielding of the institutional contacting stage of the study jointly with that for the faculty component of NSoFaS:04, the 2004 National Study of Postsecondary Faculty (NSOPF:04). A second major change was conducting student record abstraction from institutional records and student interviewing simultaneously, rather than sequentially as had been done in previous NPSAS cycles. Another change was the development of a single web-based instrument for self-administration by sample members and use by telephone interviewers alike.

We hope that the information provided in this report will be useful to interested readers. Additional information about NPSAS:04 is available on the Web at http://www.nces.ed.gov/surveys/npsas.

C. Dennis Carroll
Associate Commissioner
Postsecondary Studies Division

This page left intentionally blank.

Acknowledgments

The authors gratefully acknowledge the assistance of staff members of the National Center for Education Statistics (NCES) for their advice, guidance, and review in the design of the study, and in preparation of this document. We are particularly grateful to C. Dennis Carroll, Associate Commissioner, Postsecondary Studies Division; Paula R. Knepper, Senior Technical Advisor, Postsecondary Studies Division; James Griffith, Program Director, Postsecondary Longitudinal and Sample Survey Studies (PLSSS); Tracy Hunt-White, Statistician (PLSSS); and Linda Zimbler, Statistician (PLSSS), for their constructive input and review.

Particular thanks are also extended to the Technical Review Panel members, who provided considerable insight and guidance in the development of the design and instrumentation of the study and to Heath Brown, Melanie Corrigan, and Adam Stoll for their careful review of this report.

Thanks are also extended to the many project staff members who assisted with the production of this report. Special acknowledgment is given to John Riccobono, who served as the study's Project Director. Appreciation is extended to Jim Chromy and Roy Whitmore for their consultation and guidance on the statistical design of the study. Also, thanks to those who contributed to many of the statistical analyses presented here, including Karol Krotki, Kim Ault, Ruby Johnson, Jane Eslinger, Stephen Black, Jeremy Porter, and Darryl Creel.

Jeff Franklin, Brian Kuhr, and Robert Bozick are recognized for their contributions to the design and analysis of the data collection efforts and data quality evaluations. Finally, the tireless efforts of Sallie Fiore, Craig Hollingsworth, Sharon Powell, and Lynne Hawley in preparing this document are appreciated.

Most of all, we are greatly indebted to the institutions and students who generously participated in the survey. Their willingness to take the time to share information made this study a success.

This page left intentionally blank.

Contents

List of Tables

List of Figures

This page left intentionally blank.

Chapter 1
Overview of NPSAS:04

This document provides a description of the methodological procedures and results for the 2004 National Postsecondary Student Aid Study (NPSAS:04). The 2004 study is being conducted for the National Center for Education Statistics (NCES) of the U.S. Department of Education, Washington, DC, as authorized by Title I, Section 153 of the Education Sciences Reform Act of 2002, P.L. 107-279, 116 Stat. 1940 (2002). For reference, previous cycles of NPSAS and its longitudinal spin-off studies, the Beginning Postsecondary Students Longitudinal Study (BPS) and the Baccalaureate and Beyond Longitudinal Study (B&B), were authorized by the following legislation:

The General Education Provisions Act, as amended, 20 U.S.C. §1221 e-1 (2001).

The Higher Education Act of 1965, as amended by the Higher Education Amendments of 1986, Title XIII(a), Section 1303, and Title XIV, 20 U.S.C. §1070 et seq. (1994).

The Higher Education Act of 1965, Augustus F. Hawkins – Robert T. Stafford Elementary and Secondary School Improvement Amendments of 1988, 20 U.S.C. §2911 to 2976 (2001).

Sections 404(a), 408(a), and 408(b) of the National Education Statistics Act of 1994, 20 U.S.C. 9001 et seq. (2002).

NPSAS:04 is being conducted as the student component study of the 2004 National Study of Faculty and Students (NSoFaS:04) under contract by RTI International (RTI).[1] Results for the faculty component study of NSoFaS:04—the 2004 National Study of Postsecondary Faculty (NSOPF:04)—are provided in a separate methodology report (Heuer et al. forthcoming).

This introductory chapter describes the background, purposes, schedule, and products of the NPSAS:04 study. In chapter 2, study design and methods are described. Overall outcomes of the several stages of data collection are presented in chapter 3. Chapter 4 presents evaluations of procedures used to collect information from institutions and students and the quality of the data collected. Chapter 5 describes the procedures used in data file preparation. Chapter 6 presents the nonresponse bias analyses, weighting procedures, and variance estimation. Materials used during the study are provided as appendixes to the report and cited in the text where appropriate.

All analyses conducted to evaluate the effectiveness of the NPSAS:04 procedures are discussed. Unless otherwise indicated, a criterion probability level of 0.05 was used for all tests of significance. Throughout this document, reported numbers of sample institutions and students have been rounded to further ensure confidentiality of individual student data. As a result, row and column entries in tables may not sum to their respective totals, and reported percentages may differ somewhat from those that would result from these rounded numbers.

[1] RTI International is a trade name of Research Triangle Institute.

1.1 Background and Purpose of NPSAS

NPSAS is a comprehensive nationwide study to determine how students and their families pay for postsecondary education. The study is based on a nationally representative sample of all students (aided and nonaided) in postsecondary education institutions. Undergraduate, graduate, and first-professional students comprise the sample; these students attend all types and levels of institutions, including public and private for-profit and not-for-profit institutions, and less-than-2-year institutions to 4-year colleges and universities.

The first NPSAS study was conducted in 1986–87 to meet the need for national-level data about significant financial aid issues. Since 1987, NPSAS has been fielded every 3 to 4 years, with the last cycle conducted during the 1999–2000 academic year. Beginning in 1990, each NPSAS data collection has provided the sample and base-year data for either the BPS or the B&B. NPSAS:04 serves as the base-year study for BPS. These students will be followed up in 2006 and again in 2009.

A main objective of NPSAS:04 is to produce reliable national estimates of characteristics related to financial aid for postsecondary students. No other single national database contains student-level records for students receiving financial aid from all of the numerous and disparate programs funded by the federal government, the states, postsecondary institutions, employers, and private organizations. The data are part of NCES's comprehensive information on student financial aid and other characteristics of those enrolled in postsecondary education. The study focuses on three general questions with important policy implications for financial aid programs:

- How do students and their families finance postsecondary education?
- How does the process of financial aid work, in terms of both who applies for and who receives aid?
- What are the effects of financial aid on students and their families and on postsecondary institutions?

1.2 Major Design Changes

1.2.1 Combining NPSAS and NSOPF

For the first time, NPSAS and NSOPF were conducted together under one contract: NSoFaS:04. There has historically been a great deal of overlap in the institutional samples for these two studies since the target populations for both studies involve postsecondary institutions. To minimize institutional burden, and also to maximize efficiency in data collection procedures, the two studies were combined. This report will document the methodology and procedures used in NPSAS:04 and will discuss issues related to NSOPF when such procedures were relevant for NPSAS as well.

1.2.2 State-Representative Samples

Another important change is that NPSAS:04 was designed to provide state-level representative estimates for undergraduate students within three institutional strata—public 2-year institutions; public 4-year institutions; and private not-for-profit 4-year institutions for 12 states that were categorized into three groups based on population size—four large, four medium, and four small: California, Connecticut, Delaware, Georgia, Illinois, Indiana,

Minnesota, Nebraska, New York, Oregon, Tennessee, and Texas. These states were chosen for this "demonstration" study from a set of volunteering states that expressed interest and a willingness to support and encourage participation by their institutions.

1.3 Schedule and Products of NPSAS:04

1.3.1 Schedule

Table 1 summarizes the schedule of major activities for the full-scale study.

Table 1. Schedule of major NPSAS:04 activities: 2002–04

Activity	Start date[1]	End date[2]
Select institutional sample	8/9/02	7/18/03
Mail and make phone contact with chief administrator	3/10/03	7/17/04
Mail and make phone contact with institutional coordinator	3/24/03	7/17/04
Obtain lists for student sampling	1/7/04	7/12/04
Select student samples	1/19/04	7/13/04
Send prenotification mailing to students	2/3/04	7/22/04
Request/obtain CPS data	1/21/04	7/14/04
Preload CPS data into CADE records	1/22/04	7/20/04
Implement CADE record abstraction	2/4/04	9/9/04
Implement Web interviewing of students	2/4/04	9/9/04
Implement CATI of students	3/4/04	9/9/04

[1] This is the date on which the activity was initiated for the first applicable institution and/or its associated students.

[2] This is the date on which the activity was completed for the last applicable institution and/or its associated students.

NOTE: CPS = Central Processing System; CADE = computer-assisted data entry; CATI = computer-assisted telephone interviewing.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

1.3.2 Products

The following reports based on NPSAS:04 will be published by NCES in the future:

- *Profile of Undergraduates in U.S. Postsecondary Education Institutions: 2003–04.* Describes the demographic and enrollment characteristics of undergraduate students.
- *Student Financing of Undergraduate Education: 2003–04.* Focuses on undergraduate tuition, total price of attendance, types and sources of financial aid received, net price, financial aid need, and unmet need.
- *Student Financing of Graduate and First-Professional Education: 2003–04*. Describes the demographic and enrollment characteristics of graduate and first-professional students and the types and sources of financial aid received.

The following products have already been published and are available on the NCES website (http://nces.ed.gov/), including the first E.D. TAB and Data Analysis System (DAS):

- *2004 National Postsecondary Student Aid Study (NPSAS:04): Student Financial Aid Estimates for 2003–04*. This E.D. TAB is the first publication based on the 2003–04 data. The E.D. TAB describes the percentages of students receiving various types of financial aid and average amounts received, by type of institution attended, attendance pattern, dependency status, and income level.

- *2004 National Postsecondary Student Aid Study (NPSAS:04): Undergraduate Financial Aid Estimates for 2003–04 by Type of Institution*. This E.D. TAB is the second publication based on the 2003–04 data. This E.D. TAB focuses only on undergraduates, including separate tables for those who attended public 4-year, private-not-for-profit 4-year, public 2-year, or private for-profit postsecondary institutions during the 2003–04 academic year. It describes average tuition and fees, average total price of attendance, and the percentages of undergraduates receiving various types and combinations of financial aid and average amounts received, with a particular focus on grants and loans.
- *2004 National Postsecondary Student Aid Study (NPSAS:04): Undergraduate Data Analysis System*. The NPSAS:04 Undergraduate DAS contains the data on a sample of about 80,000 undergraduates who were enrolled at any time between July 1, 2003, and June 30, 2004, in about 1,400 postsecondary institutions. It represents all undergraduate students enrolled in postsecondary institutions in the 50 states, the District of Columbia, and Puerto Rico that were eligible to participate in the federal financial aid programs in Title IV of the Higher Education Act.
- *2004 National Postsecondary Student Aid Study (NPSAS:04): Graduate Data Analysis System*. The NPSAS:04 Graduate DAS contains the data on a sample of about 11,000 graduate students who were enrolled at any time between July 1, 2003, and June 30, 2004, in about 1,400 postsecondary institutions. It represents all graduate students enrolled in postsecondary institutions in the 50 states, the District of Columbia, and Puerto Rico that were eligible to participate in the federal financial aid programs in Title IV of the Higher Education Act.

Contact Aurora D'Amico, or visit the website (http://nces.ed.gov/dasol/) to access a NPSAS:04 DAS application or one of the NPSAS:04 reports.

Aurora D'Amico
Postsecondary Studies Division
Phone: (202) 502-7334
E-mail: aurora.d'amico@ed.gov

NPSAS:04 restricted use data files. The survey data files used to create variables in the Data Analysis Systems, and the associated electronic codebooks and file documentation, are available to researchers who have obtained a restricted data license from NCES. Information on the NCES Statistical Standards Program, including Restricted Use Data Licenses Procedures, is available from the NCES website: http://nces.ed.gov/statprog. Further information on obtaining a restricted data license may be found in the NCES *Restricted Use Data Procedures Manual* (U.S. Department of Education 1999), at http://nces.ed.gov/statprog/rudman, and also from Cynthia Barton.

Cynthia L. Barton
Data Security Assistant
Phone: (202) 502-7307
E-mail: cynthia.barton@ed.gov

Chapter 2
Design and Methodology of NPSAS:04

This chapter provides a detailed summary of the design and the methods implemented in the 2004 National Postsecondary Student Aid Study (NPSAS:04). All procedures and methods were developed in consultation with a Technical Review Panel comprised of nationally recognized experts in higher education. A complete listing of this panel is provided in appendix A. Sampling is discussed in particular detail because it occurs in several stages in this study. For example, the base-year NPSAS sample design must take into account the sampling needs for the Beginning Postsecondary Students Longitudinal Study follow-up surveys (BPS:04/06 and BPS:04/09), since the longitudinal cohort is generated from the NPSAS:04 sample. In addition, institutional contacting, instrument development, data collection procedures, data quality evaluations, and data management systems are described.

2.1 Sampling

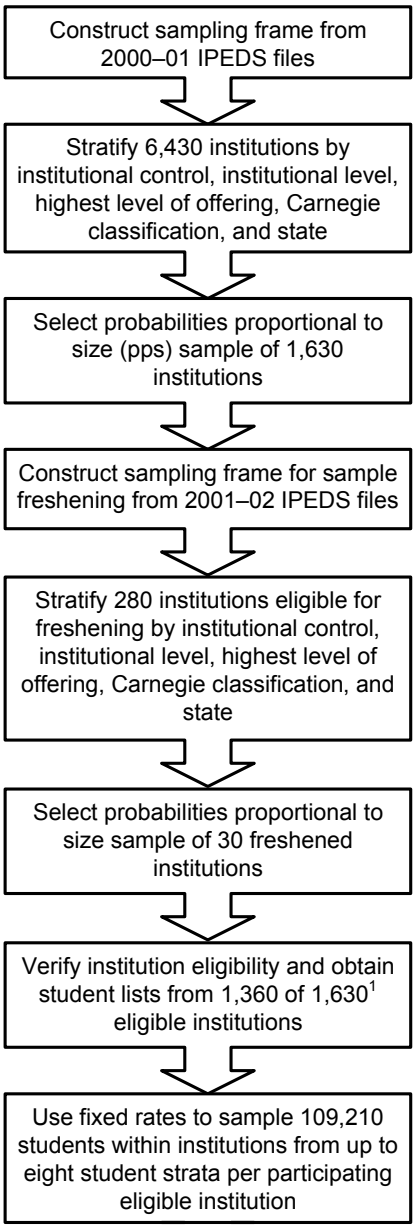
2.1.1 Target Population and Sampling Overview

The NPSAS:04 target population consists of all eligible students enrolled at any time between July 1, 2003, and June 30, 2004, in postsecondary institutions in the United States or Puerto Rico which had signed Title IV participation agreements with the U.S. Department of Education making them eligible for the federal student aid programs (Title IV institutions). To be eligible for NPSAS, students had to be enrolled in either an academic program with at least one course for credit that could be applied toward fulfilling the requirements for an academic degree or enrolled in an occupational or vocational program that requires at least 3 months or 300 clock hours of instruction to receive a degree, certificate, or other formal award. Eligible students could not be concurrently enrolled in high school and could not be enrolled solely in a general equivalency diploma (GED) or other high school completion program.

An overview of the sequential statistical sampling process for NPSAS:04 is provided in figure 1. The institution sampling frame for NPSAS:04 was constructed from the 2000–01 Integrated Postsecondary Education Data System (IPEDS) Institutional Characteristics (IC) and header files. The IPEDS data used for the initial sampling frame were collected in 2001, and the IPEDS data used for sample freshening (described in section 2.1.2) were collected in 2002. Thus, any institutions that came into existence or became eligible between the IPEDS data collections in 2002 and June 30, 2004 were not covered in the sampling frame. Institutions in the file that were not eligible (e.g., institutions located outside the United States and Puerto Rico, central offices, military academies) were deleted from the population file. The eligible institutions on the sampling frame were partitioned into 58 institutional strata based on institutional level,

institutional control, highest level of offering, Carnegie classification, and state.[2] All other students from these states were selected as part of the national sample.

Figure 1. Schematic of sequential NPSAS:04 sampling operations



[1] The 1,630 eligible institutions include the 1,630 originally selected, minus 30 ineligible institutions, plus 30 institutions from the freshened sample.

NOTE: IPEDS = Integrated Postsecondary Education Data System.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

[2] NPSAS:04 includes state-representative undergraduate student samples for three types of institutions (public 4-year, public 2-year, and private not-for-profit 4-year) in 12 states. These 12 states were selected by NCES from those expressing interest. The 12 states were categorized into three groups based on population size: four small states (Connecticut, Delaware, Nebraska, Oregon), four medium-size states (Georgia, Indiana, Minnesota, Tennessee), and four large states (California, Illinois, New York, Texas).

The 58 institutional strata, 22 nationally-representative and 36 state-representative, are shown below.

1. Public less than 2-year
2. Public 2-year, associate's Carnegie classification
3. Public 2-year, other Carnegie classification—degree-granting
4. Public 2-year, other Carnegie classification—NPSAS only
5. Public 4-year non-doctorate-granting, master's Carnegie classification
6. Public 4-year non-doctorate-granting, bachelor's Carnegie classification
7. Public 4-year non-doctorate-granting, other Carnegie classification
8. Public 4-year doctorate-granting, doctor's Carnegie classification
9. Public 4-year doctorate-granting, other Carnegie classification
10. Public 4-year NPSAS only
11. Private not-for-profit less-than-4-year, associate's Carnegie classification
12. Private not-for-profit less-than-4-year, other Carnegie classification—degree-granting
13. Private not-for-profit less-than-4-year, other Carnegie classification—NPSAS only
14. Private not-for-profit 4-year non-doctorate-granting, master's Carnegie classification
15. Private not-for-profit 4-year non-doctorate-granting, bachelor's Carnegie classification
16. Private not-for-profit 4-year non-doctorate-granting, other Carnegie classification
17. Private not-for-profit 4-year doctorate-granting, doctor's Carnegie classification
18. Private not-for-profit 4-year doctorate-granting, master's Carnegie classification
19. Private not-for-profit 4-year doctorate-granting, other Carnegie classification
20. Private not-for-profit 4-year—NPSAS only
21. Private for-profit less-than-2-year
22. Private for-profit 2-year or more
23. California public 2-year
24. California public 4-year
25. California private not-for-profit 4-year
26. Connecticut public 2-year
27. Connecticut public 4-year
28. Connecticut private not-for-profit 4-year
29. Delaware public 2-year
30. Delaware public 4-year
31. Delaware private not-for-profit 4-year
32. Georgia public 2-year
33. Georgia public 4-year
34. Georgia private not-for-profit 4-year
35. Illinois public 2-year
36. Illinois public 4-year
37. Illinois private not-for-profit 4-year
38. Indiana public 2-year
39. Indiana public 4-year
40. Indiana private not-for-profit 4-year
41. Minnesota public 2-year
42. Minnesota public 4-year
43. Minnesota private not-for-profit 4-year
44. Nebraska public 2-year
45. Nebraska public 4-year
46. Nebraska private not-for-profit 4-year
47. New York public 2-year
48. New York public 4-year
49. New York private not-for-profit 4-year
50. Oregon public 2-year
51. Oregon public 4-year
52. Oregon private not-for-profit 4-year
53. Tennessee public 2-year
54. Tennessee public 4-year
55. Tennessee private not-for-profit 4-year
56. Texas public 2-year
57. Texas public 4-year
58. Texas private not-for-profit 4-year

Institutions were selected using Chromy's sequential probability minimum replacement (pmr) sampling algorithm (Chromy 1979), which is similar to systematic sampling, to select institutions with probabilities proportional to a composite measure of size based on expected

enrollment. A sample of 1,630 institutions was selected in Fall 2002 so that these institutions could be notified early of their selection and to allow a separate sample to be selected for the field test from the remaining institutions on the sampling frame. In Summer 2003, an additional sample of about 30 institutions was selected from a frame of institutions not included on the initial sampling frame. Of the sample institutions selected for the full-scale study, about 810 were selected with certainty. The certainty institutions were either in strata in which all institutions were selected, or had expected frequencies of selection greater than unity (1.00). About 1,630 of the sampled institutions were found to be NPSAS eligible, and about 1,360 of these eligible institutions provided student enrollment lists for use as the second stage (i.e., student) sampling frame.

The sampling frames provided by sample institutions included paper and electronic lists of students enrolled in terms or courses of instruction during the previously defined NPSAS year. Student lists were sampled on a flow basis as they were received, using equal probability stratified systematic sampling. There were eight student sampling strata:

1. in-state first-time beginner students;
2. out-of-state first-time beginner students;
3. in-state other undergraduate students;
4. out-of-state other undergraduate students;
5. master's students;
6. doctoral students;
7. other graduate students; and
8. first-professional students.

First-time beginner students (FTBs) were stratified separately from other undergraduate students because they were oversampled to allow for sufficient numbers to be surveyed in the 2006 follow-up study (BPS:04/06). FTBs and other undergraduate students were each divided into in-state and out-of-state strata because undergraduate in-state students were oversampled in the 12 states with state-representative samples. These in-state and out-of-state strata were used for all institutions to allow for sampling ease and consistency; however, in states that did not have state-representative samples, in-state students were sampled at the same rate as out-of-state students.

For each student stratum, the enrollment list was sampled at a rate designed to provide approximately equal student-level probabilities. Student sampling rates were adjusted after sufficient lists had been received to accurately estimate the overall sample yield. The sampling rates were set to meet the sample sizes shown in table 2 for the national sample and table 3 for the state sample. The overall target sample size was about 121,680; however, the sampling procedures resulted in the selection of about 109,210 students. The actual sample is lower than the target sample size because institutional participation rates were somewhat lower than expected[3] and sampling rates were not adjusted high enough and early enough for the participating institutions to compensate for the loss of sample yield from the non-participating institutions.

[3] See section 3.1 for the results of institutional participation.

The sample size for NPSAS:04 is larger than past NPSAS studies. The primary reason for the increased sample size was to ensure sufficient yield for analytic purposes. The sample size was designed so that respondent yield would be sufficient for analyses even if actual response rates were lower than the targeted rates. Second, the National Center for Education Statistics (NCES) desired one weight to make the data easier for analysts to use. Also, as mentioned above, NPSAS:04 includes state-representative undergraduate student samples for three types of institutions (public 2-year, public 4-year, and private not-for-profit 4-year) in 12 states. A larger overall sample size was necessary to achieve state-representative samples in addition to the nationally-representative sample.

Table 2. Target numbers of sample students, by institutional stratum and type of student: 2004

Institutional stratum	All students	Undergraduates			Graduate students	First-professionals
		All undergraduates	FTBs	Other undergraduates		
All institutions	121,680	110,560	56,070	54,490	9,340	1,780
Public less than 2-year	4,990	4,990	4,540	440	†	†
Public 2-year	45,060	45,060	20,280	24,780	†	†
Public 4-year non-doctorate-granting	11,270	10,480	3,380	7,110	790	†
Public 4-year doctorate-granting	21,130	15,060	4,570	10,490	5,210	860
Private not-for-profit less-than-4-year	3,310	3,310	2,740	570	†	†
Private not-for-profit 4-year non-doctorate-granting	10,250	9,650	4,320	5,340	600	†
Private not-for-profit 4-year doctorate-granting	10,220	6,620	2,750	3,870	2,680	920
Private for-profit less-than-2-year	9,040	9,040	8,830	210	†	†
Private for-profit 2-year or more	6,430	6,340	4,670	1,680	80	†

† Not applicable.

NOTE: Detail may not sum to totals because of rounding. First-time beginner (FTB) and other undergraduate counts are based on the status known at the time of sampling.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Table 3. Target numbers of sample students in the 12 state representative samples, by institutional stratum and type of student: 2004

Institutional stratum	All undergraduates	FTBs	Other undergraduates
California	11,510	1,910	9,590
Public 2-year	8,620	1,120	7,500
Public 4-year	2,070	490	1,570
Private not-for-profit 4-year	820	310	520
Connecticut	1,510	660	850
Public 2-year	590	250	340
Public 4-year	500	210	290
Private not-for-profit 4-year	420	210	210
Delaware	1,770	800	970
Public 2-year	720	290	440
Public 4-year	640	320	320
Private not-for-profit 4-year	410	200	210
Georgia	2,340	1,200	1,140
Public 2-year	1,160	750	410
Public 4-year	800	280	530
Private not-for-profit 4-year	380	180	200
Illinois	4,170	1,680	2,490
Public 2-year	2,560	1,120	1,440
Public 4-year	790	230	560
Private not-for-profit 4-year	810	330	480
Indiana	1,970	910	1,060
Public 2-year	470	250	220
Public 4-year	1,010	420	600
Private not-for-profit 4-year	490	240	250
Minnesota	2,390	1,320	1,070
Public 2-year	1,360	910	440
Public 4-year	640	220	420
Private not-for-profit 4-year	390	190	200
Nebraska	1,400	650	750
Public 2-year	530	270	260
Public 4-year	580	250	330
Private not-for-profit 4-year	290	130	160
New York	5,140	2,230	2,910
Public 2-year	1,900	1,030	870
Public 4-year	1,380	410	970
Private not-for-profit 4-year	1,860	790	1,070
Oregon	1,970	860	1,110
Public 2-year	1,090	490	600
Public 4-year	590	230	360
Private not-for-profit 4-year	290	140	150
Tennessee	1,810	800	1,010
Public 2-year	750	370	380
Public 4-year	660	230	430
Private not-for-profit 4-year	400	200	200
Texas	6,260	2,970	3,290
Public 2-year	4,030	2,280	1,740
Public 4-year	1,640	450	1,190
Private not-for-profit 4-year	600	240	360

NOTE: Detail may not sum to totals because of rounding. First-time beginner (FTB) and other undergraduate counts are based on the status known at the time of sampling.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

2.1.2 Institutional Sample and Eligibility

The target population for NPSAS:04 included nearly all Title IV participating postsecondary institutions in the 50 states, the District of Columbia, and Puerto Rico.[4] To be eligible for NPSAS:04, an institution was required, during the 2003–04 academic year, to

- offer an educational program designed for persons who had completed secondary education;
- offer at least one academic, occupational, or vocational program of study lasting at least 3 months or 300 clock hours;
- offer courses that were open to more than the employees or members of the company or group (e.g., union) that administered the institution;
- be located in the 50 states, the District of Columbia, or Puerto Rico;
- be other than a U.S. Service Academy;[5] and
- have a signed Title IV participation agreement with the U.S. Department of Education.

As indicated above, institutions providing only avocational, recreational, or remedial courses or only in-house courses for their own employees were excluded. The listed eligibility requirements are consistent with those used in previous NPSAS rounds, with two exceptions: the last requirement was new for NPSAS:2000, and offering more than just correspondence courses was no longer a requirement beginning with NPSAS:04.

The student sample was allocated to the separate applicable institutional and student sampling strata, defined above. Student sampling rates, which were used to compute institution-level composite measures of size, were based on the 2000 IPEDS Fall Enrollment Survey counts and the required sample sizes (see appendix B for details).

An independent sample of institutions was selected for each institutional stratum using Chromy's sequential probability minimum replacement (pmr) sampling algorithm (Chromy 1979) to select institutions with probabilities proportional to their computed measures of size. However, rather than multiple selections of sample institutions being allowed,[6] those with expected frequencies of selection greater than unity (1.00) were selected with certainty. The remainder of the institutional sample was selected from the remaining institutions within each stratum. The sampling algorithm was implemented with a random start for each institutional stratum to ensure the positive pairwise probabilities of selection that were needed for proper variance estimation (Chromy 1981).

The sample of institutions was initially selected in September 2002 to allow the field test sample institutions to be selected from the complement of the full-scale sample. In July 2003, a freshened sample of institutions was selected from a frame of institutions that were not on the

[4] Title IV participating institutions excluded from the target population were the five U.S. Service Academies.
[5] These academies were not eligible for this financial aid study because of their unique funding/tuition base.
[6] Precluding institutions with multiple selections at the first stage of sampling made it unnecessary to select multiple second-stage samples of students.

original sampling frame because they were either new institutions or newly eligible institutions.[7] Freshening was done to ensure the representativeness of the sample because the initial sample was selected a year earlier. The measures of size for the supplemental sampling frame from which the freshened sample was selected were based on the 2002 IPEDS Fall Enrollment Survey counts.

Table 4 shows the institution sampling rates and the numbers of certainty and noncertainty institutions selected for each of the 22 national strata and the 36 state strata, respectively. The institutions included in the national sample were selected from all 58 strata, while institutions included in the state samples were selected only from the 36 state strata. Within each institutional stratum, additional implicit stratification was accomplished by sorting the stratum sampling frame by the following classifications: (1) historically Black colleges and universities (HBCU) indicator; (2) Carnegie classifications of postsecondary institutions; (3) the Office of Business Economics (OBE) Region from the IPEDS header file (Bureau of Economic Analysis of the U.S. Department of Commerce Region);[8] and (4) the institution measure of size. The objective of this implicit stratification was to approximate proportional representation of institutions on these measures.

Table 4. Institutional sampling rates and number of certainty and noncertainty institutions, by institutional stratum: 2004

Institutional stratum[1]	Size of universe[2]	Sampling rate	Number of sample institutions		
			Total	Certainty	Noncertainty
Total	6,706	0.25	1,670	810	860
Public less than 2-year	317	0.21	70	20	50
Public 2-year associate	623	0.12	70	#	70
Public 2-year other—degree-granting	36	0.14	10	#	#
Public 2-year other—NPSAS only[3]	69	0.45	30	10	20
Public 4-year non-doctorate-granting, master's	118	0.17	20	#	20
Public 4-year non-doctorate-granting, bachelor's	65	0.17	10	#	10
Public 4-year non-doctorate-granting, other	47	0.06	#	#	#
Public 4-year doctorate-granting, doctor's	126	1.00	130	130	#
Public 4-year doctorate-granting, other	49	0.20	10	#	10
Public 4-year NPSAS only[3]	16	0.13	#	#	#
Private not-for-profit less-than-4-year, associate	108	0.31	30	#	30
Private not-for-profit less-than-4-year, other—degree-granting	24	0.08	#	#	#
Private not-for-profit less-than-4-year, other—NPSAS only[3]	240	0.16	40	10	30
Private not-for-profit 4-year non-doctorate-granting, master's	132	0.09	10	#	10
Private not-for-profit 4-year non-doctorate-granting, bachelor's	293	0.12	30	#	30
Private not-for-profit 4-year non-doctorate-granting, other	202	0.16	30	#	30
Private not-for-profit 4-year doctorate-granting, doctor's	52	1.00	50	50	#
Private not-for-profit 4-year doctorate-granting, master's	61	0.18	10	#	10
Private not-for-profit 4-year doctorate-granting, other	143	0.09	10	#	10
Private not-for-profit 4-year—NPSAS only[3]	51	0.06	#	#	#
Private for-profit less-than-2-year	1,445	0.12	170	10	170
Private for-profit 2-year or more	1,149	0.10	110	10	110

See notes at end of table.

[7] Some of the IPEDS data provided by institutions that was used to determine eligibility for the original frame was sufficiently different from the IPEDS data subsequently provided by institutions to determine eligibility for the freshening frame.

[8] For sorting purposes, Alaska and Hawaii were combined with Puerto Rico in the Outlying Areas region rather than in the Far West region.

Table 4. Institutional sampling rates and number of certainty and noncertainty institutions, by institutional stratum: 2004—Continued

Institutional stratum[1]	Size of universe[2]	Sampling rate	Number of sample institutions		
			Total	Certainty	Noncertainty
California	298	0.38	110	50	60
Public 2-year	114	0.33	40	#	40
Public 4-year	33	1.00	30	30	#
Private not-for-profit 4-year	151	0.27	40	20	20
Connecticut	45	1.00	50	50	#
Public 2-year	15	1.00	20	20	#
Public 4-year	10	1.00	10	10	#
Private not-for-profit 4-year	20	1.00	20	20	#
Delaware	9	1.00	10	10	#
Public 2-year	3	1.00	#	#	#
Public 4-year	2	1.00	#	#	#
Private not-for-profit 4-year	4	1.00	#	#	#
Georgia	108	0.79	90	60	30
Public 2-year	53	0.57	30	10	30
Public 4-year	21	1.00	20	20	#
Private not-for-profit 4-year	34	1.00	30	30	#
Illinois	148	0.49	70	40	40
Public 2-year	48	0.63	30	10	20
Public 4-year	12	1.00	10	10	#
Private not-for-profit 4-year	88	0.34	30	10	20
Indiana	71	0.85	60	50	10
Public 2-year	16	1.00	20	20	#
Public 4-year	14	1.00	10	10	#
Private not-for-profit 4-year	41	0.73	30	20	10
Minnesota	90	0.86	80	70	10
Public 2-year	43	0.70	30	20	10
Public 4-year	11	1.00	10	10	#
Private not-for-profit 4-year	36	1.00	40	40	#
Nebraska	29	1.00	30	30	#
Public 2-year	7	1.00	10	10	#
Public 4-year	7	1.00	10	10	#
Private not-for-profit 4-year	15	1.00	20	20	#
New York	249	0.43	110	70	30
Public 2-year	37	1.00	40	40	#
Public 4-year	45	0.67	30	20	10
Private not-for-profit 4-year	167	0.24	40	20	20
Oregon	52	1.00	50	50	#
Public 2-year	17	1.00	20	20	#
Public 4-year	10	1.00	10	10	#
Private not-for-profit 4-year	25	1.00	30	30	#
Tennessee	75	0.81	60	50	10
Public 2-year	21	1.00	20	20	#
Public 4-year	10	1.00	10	10	#
Private not-for-profit 4-year	44	0.68	30	20	10

See notes at end of table.

Table 4. Institutional sampling rates and number of certainty and noncertainty institutions, by institutional stratum: 2004—Continued

Institutional stratum[1]	Size of universe[2]	Sampling rate	Number of sample institutions		
			Total	Certainty	Noncertainty
Texas	166	0.54	90	50	40
Public 2-year	68	0.44	30	10	20
Public 4-year	43	0.70	30	20	10
Private not-for-profit 4-year	55	0.55	30	20	10

Rounds to zero.

[1] Stratum reflects institutional categorization as determined from the 2000–01 Integrated Postsecondary Education Data System (IPEDS) file; some institutions were categorized differently in later IPEDS files.

[2] Based on the 2000–01 and 2002–03 IPEDS files.

[3] "NPSAS-only" refers to institutions that were not included on the sampling frame for NSOPF—the faculty component of NSoFaS.

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

2.1.3 Student Sample and Eligibility

The postsecondary students eligible for NPSAS:04 were those who attended a NPSAS-eligible institution during the 2003–04 academic year and who were

- enrolled in *either* (1) an academic program; (2) at least one course for credit that could be applied toward fulfilling the requirements for an academic degree; *or* (3) an occupational or vocational program that required at least 3 months or 300 clock hours of instruction to receive a degree, certificate, or other formal award;
- not concurrently enrolled in high school; and
- not enrolled *solely* in a GED or other high school completion program.

Each sampled institution that was verified as NPSAS-eligible was asked to provide a list of all its students who satisfied all the NPSAS eligibility conditions, preferably an "unduplicated" electronic list (i.e., one in which each student's name appeared only once), together with identifying, classifying, and locating information (see section 2.3.2). Although electronic files were preferred, student lists were accepted in a variety of formats, as long as they were complete.

Several checks on quality and completeness of student lists were implemented before the sample students were selected. Institutions providing lists that failed these checks were contacted to resolve the detected problems. Enrollment lists failed quality control checks under the following conditions:

- FTBs were not identified (unless the institution only enrolled graduate/first-professional students or explicitly indicated that no FTBs existed in the school); and/or
- student level—undergraduate, master's, doctoral, other graduate, or first professional—was not clearly identified.

Quality checks on student counts were performed separately for FTBs and all other students. The "unduplicated" FTB counts were checked against the fall enrollment counts from the IPEDS Fall Enrollment Survey because IPEDS does not have "unduplicated" annual FTB

counts. The check failed if the count for any "unduplicated" list was at least 50 percent less than the IPEDS count. The list counts were expected to almost always be more than the IPEDS counts because the IPEDS counts were not annual counts. This check identified institutional enrollment lists that under-reported FTBs. The "unduplicated" counts of other undergraduates, graduates, and first-professionals were checked against the "unduplicated" annual enrollment counts from the IPEDS Fall Enrollment Survey. The check failed if the count for any "unduplicated" list differed by at least 50 percent from the IPEDS count.[9]

As student lists were received from institutions, students were sampled using predetermined sampling rates that varied by student stratum. Stratified systematic sampling was used to ensure comparable sampling procedures for both paper and electronic lists. After the sample of students had been selected for an institution, Social Security numbers (SSNs) of those sampled were compared to those of students who had already been selected from other institutions to eliminate cross-institution duplication. Multiplicity adjustments in the sample weighting (described in more detail in section 6.2.1) accounted for the fact that any students who attended more than one institution during the NPSAS year had more than one chance of selection.

Some institutional systems sent in lists for multiple institutions or campuses. If the lists were separate for each institution or campus, then the samples were selected separately and independently. If the lists were combined into one list with no identifier mapping students to institution or campus, then one student sample was selected that represented all of the institutions or campuses included on the list. In such cases, sampling rates were adjusted, and a weight adjustment was made (see section 6.1.1).

For paper lists, samples were selected manually, and then the list of sample students was entered into an electronic file. When students from different strata (e.g., FTBs and other undergraduates) were combined on a paper list, the sampling rate from the stratum with the higher rate was used. Then after the sample was entered into an electronic file, the students from the other stratum (or strata) were subsampled to match the sampling rates for that stratum.[10]

Initial student sampling rates were calculated for each sample institution using sampling rates designed to generate approximately equal probabilities of selection within the ultimate institution-by-student sampling strata (see appendix B). However, these rates were sometimes modified as follows:

- Student sampling rates were increased, as needed, so that the sample size achieved at each sample institution would be at least 10 sample students, where possible, to ensure sufficient yield for variance estimation.
- Student sampling rates were decreased if the sample size was more than 50 greater than the institution had been told to expect, which was based on the sampling rate applied to the enrollment count on the sampling frame.[11]

[9] If provided paper lists were not "unduplicated," an "unduplicated" total was estimated by applying an empirically determined multiplicity factor (0.50) to the student count from the provided lists.
[10] The issue of combined strata was not a problem for electronic lists since the file could be sorted by stratum prior to sampling.
[11] This was to ensure minimal burden for the institutions participating in computer-assisted data entry (CADE) data abstraction.

- Sample yield was monitored throughout enrollment list collection and student sampling rates were adjusted periodically for institutions for which sample selection had not yet been performed to ensure that the desired student sample sizes were achieved.

These adjustments to the initial sampling rates resulted in some additional variability in the student sampling rates and, hence, in some increase in survey design effects (variance inflation—see section 6.4.3).

The planned and achieved sample sizes by student stratum and level of offering are shown in table 5. The initial classification of the student sample overall and by institution type and student stratum are shown in table 6. As mentioned earlier, the achieved sample yield was less than what was planned (109,210 students as compared to the target of 121,680). Institutional participation rates were somewhat lower than expected, and sampling rates were not adjusted high enough and early enough for the participating institutions to compensate for the loss of sample yield from the non-participating institutions. Overall, there were more doctoral and other graduate students in the sample than planned, and there were fewer FTBs, other undergraduate students, and master's students than planned. (See appendix B, section B.4 for additional detail on the sample allocation.)

Table 5. Planned and achieved NPSAS:04 student samples, by student stratum and level of offering: 2004

Student stratum[1]	Institutional level[2]	Students sampled		
		Number expected[3]	Number achieved[4]	Percent[5]
Total	All institutions	121,680	109,210	89.8
FTB	Subtotal	56,070	49,410	88.1
	Less-than-2-year	14,080	11,370	80.8
	2- to 3-year	24,530	22,250	90.7
	4-year	42,700	15,790	37.0
Other undergraduate	Subtotal	54,490	47,680	87.5
	Less-than-2-year	800	920	115.1
	2- to 3-year	25,990	19,660	75.6
	4-year	27,690	27,100	97.9
Master's	4-year	5,310	3,720	70.1
Doctor's	4-year	3,630	4,950	136.1
Other graduate	4-year	400	1,660	416.3
First-professional	4-year	1,780	1,790	100.7

[1] As expected the sampling frames misclassified some individual students with respect to first-time beginner (FTB), undergraduate, graduate, and first-professional status; statistics presented in this table are based on the sampling frame classification. The two FTB strata (in-state and out-of-state) have been combined, and the two other undergraduate strata (in-state and out-of-state) have been combined.

[2] Institutional level is based on the 2003–04 Integrated Postsecondary Education Data System (IPEDS) file. This file was used to reflect the level during the NPSAS year, which may be different than the level at the time of sampling.

[3] Based on sample allocation.

[4] The student sample was drawn from 1,360 eligible institutions that provided enrollment lists.

[5] Percent reported reflects the ratio of "achieved" to "expected."

NOTE: Detail may not sum to totals because of rounding. FTB = first-time beginner.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Table 6. Initial classification of NPSAS:04 student sample, by institutional characteristics and student stratum

	Total sample[1]		Student sampling stratum[2]							
			FTB sample[3]		Other undergraduate sample		Graduate sample[3]		First-professional sample	
Institutional characteristics	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
All institutions	109,210	100.0	49,410	100.0	47,680	100.0	10,330	100.0	1,790	100.0
Institutional level										
Less-than-2-year	12,310	11.3	11,370	23.0	920	1.9	20	0.1	#	0.2
2-year	41,960	38.4	22,250	45.0	19,660	41.2	40	0.4	10	0.5
4-year non-doctorate-granting	21,550	19.7	8,220	16.6	12,130	25.4	1,180	11.4	20	0.9
4-year doctorate-granting	33,400	30.6	7,570	15.3	14,970	31.4	9,100	88.1	1,770	98.4
Institutional control										
Public	71,030	65.0	27,820	56.3	35,720	74.9	6,570	63.6	920	51.3
Private not-for-profit	22,730	20.8	8,770	17.8	9,450	19.8	3,640	35.2	870	48.4
Private for-profit	15,460	14.2	12,820	26.0	2,510	5.3	120	1.2	#	0.2
Type of institution										
Public less-than-2-year	2,780	2.5	2,330	4.7	440	0.9	#	#	#	#
Public 2-year	36,340	33.3	17,780	36.0	18,520	38.8	30	0.3	10	0.5
Public 4-year non-doctorate-granting	9,210	8.4	2,680	5.4	5,970	12.5	550	5.4	10	0.4
Public 4-year doctorate-granting	22,700	20.8	5,030	10.2	10,790	22.6	5,980	57.9	900	50.4
Private not-for-profit 2-year or less	3,020	2.8	2,350	4.8	670	1.4	10	#	#	#
Private not-for-profit 4-year non-doctorate-granting	9,310	8.5	3,920	7.9	4,840	10.2	540	5.3	10	0.4
Private not-for-profit 4-year doctorate-granting	10,400	9.5	2,510	5.1	3,940	8.3	3,090	29.9	860	48.0
Private for-profit less-than-2-year	8,750	8.0	8,280	16.8	460	1.0	10	0.1	#	0.2
Private for-profit 2-year or more	6,710	6.1	4,540	9.2	2,050	4.3	110	1.1	#	0.1

Rounds to zero.
[1] The student sample was drawn from 1,360 eligible institutions that provided enrollment lists.
[2] As expected, the sampling frames misclassified some individual students as to first-time beginner (FTB), undergraduate, graduate, and first-professional status; statistics presented in this table are based on the sampling frame classification. This explains why some graduate/first-professional students were sampled from institutions that do not have such students.
[3] The two FTB strata (in-state and out-of-state) have been combined, the two other undergraduate strata (in-state and out-of-state) have been combined, and the master's, doctorate, and other graduate strata have been combined.

NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

2.2 Sources of Data

Information for NPSAS:04 was obtained from several sources, including the following:

- **Student Record abstraction (computer-assisted data entry [CADE]):** Data from institutional financial aid and registrar records at the sampled institutions currently attended. These data were entered at the institution by institutional personnel or field data collectors in 2003–04 using a web-based computer-assisted data entry program (web-CADE) or directly downloaded to a data file (data-CADE).
- **Student Interview:** Data collected directly from sampled students via web-based self-administered or interviewer-administered questionnaires.
- **Central Processing System (CPS):** U.S. Department of Education database of federal financial aid applications for the 2003–04 academic year. Data provided by students on the Free Application for Federal Student Aid (FAFSA) form.
- **National Student Loan Data System (NSLDS):** U.S. Department of Education database of federal Title IV loans and Pell Grants. The accessed NSLDS Pell Grant and loan files included information for the year of interest, as well as a complete federal grant or loan history for each applicable student.
- **Integrated Postsecondary Education Data System (IPEDS):** U.S. Department of Education, National Center for Education Statistics, database of descriptive information about individual postsecondary institutions attended by sample students.

These diverse and sometimes overlapping data sources provided some information that could not be collected directly from institutions or students. They also provided a way to "fill in" certain data that were also gathered via student record abstraction or the student interview but were missing for individual sample members (e.g., demographics). Finally, these overlapping data sources sometimes served to check or confirm the accuracy of similar information obtained from other sources.

2.3 Data Collection Design

As mentioned in the previous section, NPSAS data are gathered from multiple sources, some directly from institutions and students, and some from extant data sources. The various data collections will be described in the following sections. As with previous rounds of NPSAS, the first step involved contacting the institutions, describing the nature and purpose of the study, identifying institutional coordinators, and asking for institutional participation. Next, institutions were asked to provide lists of enrolled students from which the student sample could be selected. Student-level data were then collected via the institutional student record abstraction and the student interview.

Two important changes of note involve the sequence of student-level data collection processes. In past rounds of NPSAS, institutions were not asked to provide any contact information for students until the student sample had been selected. Information needed to locate and contact students for participation in the student interview was collected as part of the student record abstraction, to avoid unduly burdening institutions by asking for information for students that would not ultimately become part of the student sample. However, in the past, the sequential

linkage between CADE record abstraction and the student interview has adversely impacted the overall data collection schedule, and in turn, subsequent release of the data. Therefore, in NPSAS:04, student contact information was obtained with the enrollment lists, so that student interviewing could occur simultaneously with CADE and, thereby, reduce the amount of time required for data collection.

Another significant change in data collection procedures was the introduction of a single web-based instrument for both self-administered and interviewer-administered student interviews, which benefited the study in several ways, including facilitating the expeditious processing and documentation of data files.

The following sections describe the procedures implemented at each stage of data collection in more detail.

2.3.1 Institutional Contacting

Training

Three training sessions were held for institution contactors. In each session, institution contactors were trained to

- prompt institutions to provide requested data within schedule constraints;
- handle help desk questions on all components;
- avert and convert refusals;
- deal effectively with gatekeepers and other institutional staff; and
- use the Institutional Contacting System[12] (ICS) to document calls, schedule appointments, and send problems to project staff for resolution.

The first training session focused on institution recruitment—contacting the office of the chief administrator, making an initial contact to the designated institution coordinator, and prompting for completion of the Coordinator Response Form. The second training coincided with the mailing of the complete the National Study of Faculty and Students (NSoFaS) binder to the coordinators, and focused on prompting for student and faculty lists. The third training included an introduction to the CADE component, and focused on coordinating data collection and prompting activities for the student and faculty components of NSoFaS.

Each training session consisted of 2 days of classroom instruction and practice sessions in which contactors paired off with other contactors to rehearse prompting calls, answering help desk questions and using the ICS.

Additional ad hoc trainings on specific issues (refusal aversion and conversion, handling multi-campus institutions, etc.) were held as needed, often as part of regularly scheduled quality control meetings.

[12] The ICS is designed to track and document the status of sample institutions through the various phases of the project including initial contacting, coordinator contracting, enrollment list preparation, sampling, and data collection. See section 2.4.2 for more detail.

Institutional contacting

The eligible institutional sample for NSoFaS:04 consisted of about 1,630 institutions, all of which were sampled for NPSAS and 1,080 of which were also sampled for the National Study of Postsecondary Faculty (NSOPF). The process of recruiting institutions and initiating coordinator contacts began well before the beginning of the academic year of interest for several reasons. First, such early notification allowed schools time to plan for the resources required for participation within the study's schedule constraints. Early contacting also allowed institutions enough time for any required internal review and approval procedures, and time for institutions to work with project staff to resolve any potential obstacles to their participation. This advance notification was intended to increase the institutional response rate, accelerate the receipt of student lists, and increase the response rate of student sample identification.

Prior to the field test, endorsements from major professional associations and organizations that had previously endorsed NPSAS were renewed, as appropriate, to both NSoFaS component studies. An effort was also made to solicit new endorsements from other organizations. In all, 25 organizations endorsed NSoFaS.[13] These endorsements were featured on all project letterhead, pamphlets, and on the NSoFaS website. In addition, several of these organizations continued to promote the study throughout the data collection period in newsletters and other communications.

For NPSAS, the overall process of student enrollment list collection proceeded according to the following steps which are described in detail below:

- initial contact;
- institution recruitment, and
- student list collection.

Initial contact. Institution contactors were hired and initially trained to confirm the name and contact information for the chief administrator, who served to confirm the institution's intention to participate in the study. Institutional eligibility was also confirmed at this time.

Institutions flagged as potentially ineligible—including closed institutions and institutions that indicated they were not Title IV eligible or open to the general public—were reviewed by project staff. Instances of sampled institutions that merged with other institutions (sampled or unsampled), possible changes in mission that could affect the institution's sampling strata, and changes in name or address were also reviewed.

Institution recruitment

Notification materials. Institution recruitment began in Spring 2003. Chief administrators at institutions sampled for NSoFaS were sent the following materials. (Copies of letters and pamphlets sent to chief administrators and institutional coordinators can be found in appendix C.)

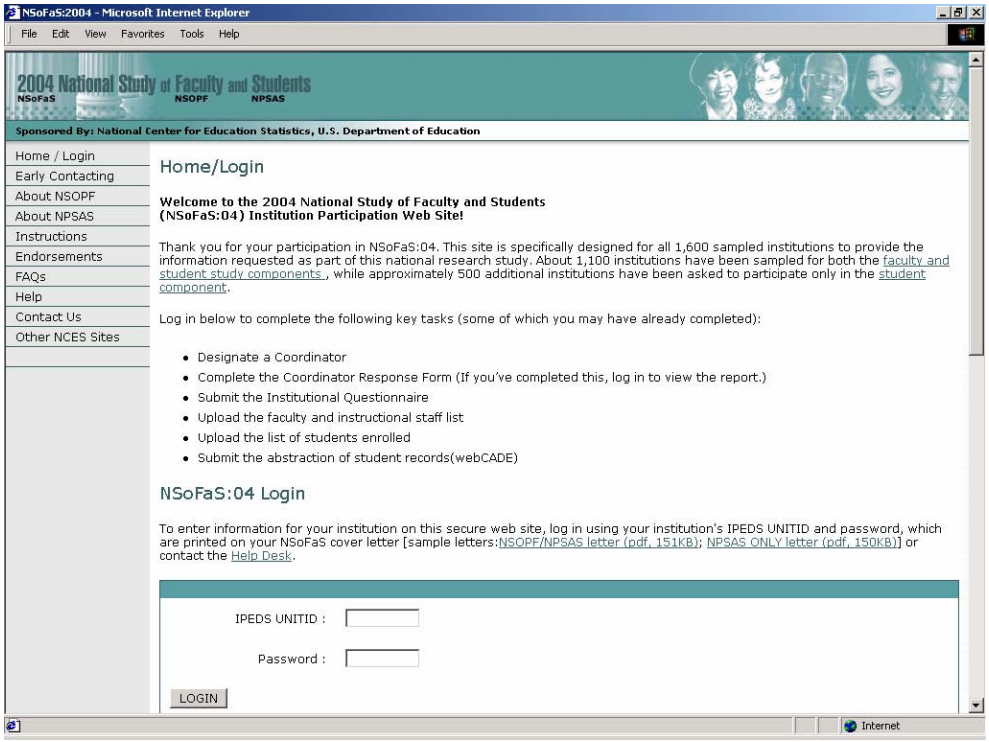
- A cover letter, printed on NCES letterhead, providing background information on NPSAS and NSOPF.[14] The letter requested that the chief administrator designate an institutional coordinator.

[13] One of these organizations, associated with for-profit schools, was asked only for an endorsement for NPSAS.
[14] Materials regarding NSOPF were included only to institutions that were also selected to participate in NSOPF.

- An NSoFaS pamphlet summarized the objectives of both NPSAS and NSOPF, and provided background information and selected findings for each component.[15]
- A NPSAS pamphlet, included to show what had been prepared for sampled students.
- A project timeline outlining the flow of activities for both component studies of NSoFaS, and the projected schedule for each.
- If sampled for NSOPF, an NSOPF pamphlet was included to show what had been prepared for mailing to the sampled faculty.

Institution website. A website was developed for use by institutions selected for participation in NSoFaS and the address was provided in all materials sent to institutions. The NSoFaS website served a number of functions for institutions selected for participation in NSoFaS. In addition to providing general information about the NPSAS and NSOPF studies being conducted, it served as a central repository for all study documents and instructions. It also allowed for the uploading of electronic lists of enrolled students. Figure 2 presents the home page of the NSoFaS website.

Figure 2. The 2004 National Study of Faculty and Students institution website home page



SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Study of Faculty and Students (NSoFaS:04) website.

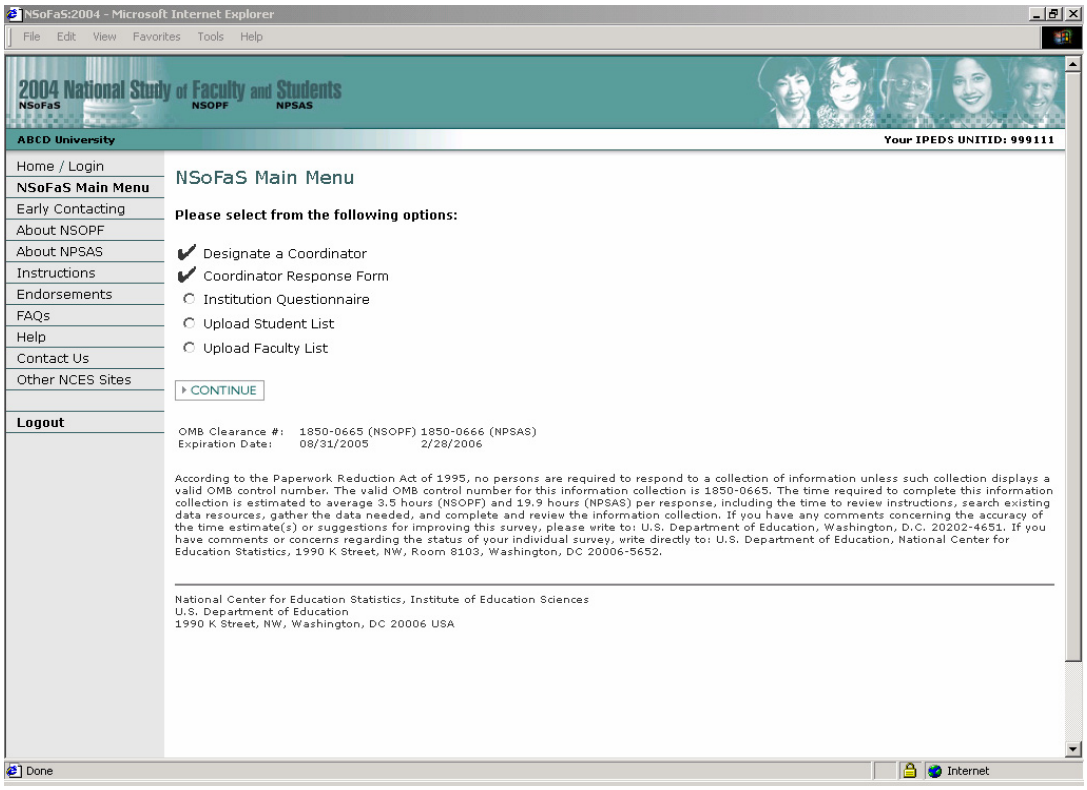
[15] The institution website provided all necessary information and documentation to institutions that participated in NSoFaS. All were selected for NPSAS and many were also selected for NSOPF. Study-specific materials were provided as appropriate.

Visitors to the website were provided with the following links (see navigation bar on the left side of the screen):

- *Early Contacting* provided information about the early institution contacting for NSoFaS:04 for the initial stage.
- *About NPSAS* and *About NSoPF* provided information on each study's mandate and research objectives, with a link to NCES reports from previous study cycles.
- *Instructions* provided links that allowed institution staff to view and print copies of various NPSAS and NSoPF forms.
- *Endorsements* listed the 25 national organizations that endorsed the studies. (These are listed in appendix D.)
- *Frequently Asked Questions (FAQs)* included questions and answers concerning all stages of data collection for both components of NSoFaS.
- *Help* provided the help desk toll-free number and e-mail address for contacting project staff, along with instructions for logging in.
- *Contact Us* contained address information for RTI.
- *Other NCES Sites* linked to three NCES web pages that provided more information about NCES programs:
  - Site map of NCES website—http://nces.ed.gov/help/sitemap.asp;
  - Postsecondary Education Studies—http://nces.ed.gov/surveys/surveygroups.asp?group=2; and
  - To order publications and products—http://nces.ed.gov/pubsearch.

A status screen, shown in figure 3, indicated which stages of institution data collection were completed (denoted by a check mark) and allowed institutions to select from those stages that were not yet completed. Once a stage was completed, it was no longer accessible via the Web.

Figure 3. The 2004 National Study of Faculty and Students (NSoFaS:04) institution website status screen



Source: U.S. Department of Education, National Center for Education Statistics, 2004 National Study of Faculty and Students (NSoFaS:04) website.

Designation of institutional coordinator. A team of institutional contactors followed up with the chief administrators by telephone. The chief administrators were asked to name an institutional coordinator whose role was to respond to requests for data and coordinate data production and delivery efforts. Once an institutional coordinator was designated, they received the same packet of notification materials described above.

Working with Institutional Review Boards. Institutional coordinators who indicated that a formal review process, such as an Institutional Review Board (IRB) review, was necessary before their institution would agree to participate were forwarded additional project materials as appropriate. A complete IRB packet was prepared for this purpose and mailed to the coordinator upon request. This packet included copies of questionnaires, as well as complete descriptions of relevant survey procedures, including confidentiality and informed consent.

2.3.2 Student Enrollment List Acquisition

Complete instructions for providing the student enrollment lists, and other requested materials were provided to institutional coordinators.

Due dates for providing the enrollment list of students requested for NPSAS were based on the term structure of each institution. Institutions were encouraged to submit an electronic list by uploading it to the secure website. The data items requested for each listed student were the following:

- full name;
- student ID;
- Social Security number;
- educational level;
- FTB status (defined as one with no transfer credits from another institution, first enrolled as a freshman between July 1, 2003, and April 30, 2004, or has not completed a postsecondary class prior to July 1, 2003);
- local address;
- local telephone number;
- campus e-mail;
- permanent address; and
- permanent e-mail.

Follow-up with institutional coordinators was conducted by telephone, mail, and e-mail. Telephone prompts to the institutional coordinators were made for institutions that had not provided lists. E-mail reminders that encouraged participation were sent to institutional coordinators prior to pending deadlines. As enrollment lists were received, they were reviewed for completeness, readability, and accuracy. Additional follow-up to clarify the information provided or retrieve key missing information was conducted by the institution contactors as necessary. This included follow-up with institutions that failed quality control checks against IPEDS files, and institutions that failed to provide key variables (FTB status, etc.).

Reimbursement for staff time involved in providing student lists was offered to institutions reporting difficulty meeting the schedule for submitting lists. A refusal conversion letter was mailed to institutions that had not responded.

Systemwide participation and multi-campus enrollment lists

In some instances, state postsecondary systems and private multi-campus institutions were able to provide enrollment lists for all their sampled institutions from a central office. In these instances, a "lead institution" was appointed, and a coordinator was designated to report for all sampled institutions.

Systemwide offices also provided other data collection assistance. One large multi-campus system devised a software program that would allow institutions within the system to easily download the information requested for the list in a usable format and distributed the software to their sampled campuses. Others—particularly within the 12 oversample states—actively encouraged their campuses to participate. More than 200 institutions reported as part of a multi-campus system.

Student enrollment lists from NSLDS

To increase representation within certain strata in which institutional participation was low, some student lists were obtained directly from NSLDS records for individual institutions, rather than the institutions themselves. These lists had two important drawbacks which limited

their usefulness to a small number of institutions for which reliable lists could not otherwise be obtained. First, NSLDS lists only contained records for federal financial aid recipients, and did not represent all enrolled students. Second, the NSLDS lists did not contain as much locating data for students as did enrollment lists provided by institutions. Thus, additional locating information had to be obtained to contact the students. For these reasons, NSLDS lists were used only when most students at these institutions were thought to be aid recipients.[16] NSLDS lists were used for sampling for only about 10 institutions. Among these 10 institutions for which sampling frames were obtained from NSLDS, 55 percent were for-profit less than 2-year institutions, 36 percent were for-profit, and 9 percent were private not-for-profit less than 4-year institutions.

12-state cooperation and assistance

A point of contact was identified in each of the 12 states with representative samples of undergraduates at the state level. These individuals were regularly updated on the participation status of institutions within their states. They also assisted with ongoing efforts to encourage institutional participation by contacting the chief administrators and institution coordinators at sampled institutions.

2.3.3 Matching to Federal Databases (CPS, NSLDS)

To reduce institutional burden in subsequent study data collections, information related to applications for federal financial aid during the financial aid year was obtained from the U.S. Department of Education's CPS. Students enter this information on the FAFSA form; it is then converted to an electronic form, analyzed, and provided to requesting institutions and other approved parties. As was the case in NPSAS:96 and NPSAS:2000, RTI was assigned a "special designation code" by CPS. Under this procedure, financial aid application data were requested through a standard Federal Data Request process.[17] The CPS was accessed daily to download data from the completed request.

Data on the nature and amount of Pell Grants or federal student loans were obtained from the NSLDS database maintained by the U.S. Department of Education. The electronic data interchange with NSLDS was performed twice during the data collection period and once after data collection ended in order to send the most up-to-date data for matching as possible. It included a query of both federal student loan and Pell Grant files. A successful match with the NSLDS loan and Pell database required that the student have a valid application record within the database. The accessed NSLDS Pell Grant and loan files included information for the year of interest, as well as a complete federal grant or loan history for each applicable student.

[16] Student enrollment lists were used from NSLDS when IPEDS data indicated that the percentage of grant-receiving students was at least 80 percent, and the percentage of students receiving loans was at least 90 percent. In most cases, both percentages were higher than 90 percent.

[17] This is a request process similar to that available to state and federal requests from the system, through which information can be requested about individuals regardless of the institution they attend. Requests made by an institution are restricted to applicants to that institution only.

2.3.4 Data Abstraction from Student Records (CADE)

Instrument development

Three modes were used for student record abstraction: 1) institutions entered data directly into the web-based CADE system (referred to as self-CADE); 2) institutions provided student record information in data files according to specifications (data-CADE); and 3) trained RTI field data collectors abstracted the student record data into the web-based CADE system (field-CADE). The web-based CADE system was created using Active Server Pages technology against a structured query language (SQL) server database. The overall content of the NPSAS:04 CADE instrument was very similar to the instrument used in NPSAS:2000 and NPSAS:96 as it had worked very well in obtaining the desired data elements from the institutions. However, the instrument was modified so that NPSAS:2000 items specific to the B&B cohort were deleted and items necessary to identify the BPS cohort were added.

A facsimile of the CADE instrument is presented in appendix E. It consisted of three sections grouped by topic. The first section collected financial aid information and included three subsections: financial aid awards, need analysis, and Institutional Student Information Report (ISIR). The second section collected registration and admissions information and it also contained three subsections: locating, student characteristics, and admissions tests. The third and last section consisted of two subsections: enrollment and tuition. Figure 4 shows the layout of the CADE instrument along with additional details from each subsection.

Figure 4. Structure and content of computer-assisted data entry (CADE) student record abstraction instrument: 2004

Section 1: Financial Aid Information

- Financial Aid
  - Awards and associated amounts, categorized by source of award
- Need Analysis
  - Expected family contribution (EFC)
  - Dependency status
  - Cost of attendance/budget
- Institution Student Information Record (ISIR)
  - Social Security number and last name from ISIR for purposes of matching Central Processing System (CPS)

Section 2: Registration and Admissions

- Locating
  - Local and permanent addresses, phone numbers, and e-mail addresses
- Characteristics
  - Date of birth
  - Gender
  - Marital status
  - High school degree
  - Race
- Admissions Tests
  - Scholastic Aptitude Test (SAT) and American College Testing (ACT) scores for undergraduates
  - Graduate Record Exam (GRE) scores for graduate and first professional students

Section 3: Enrollment and Tuition

- Enrollment
  - Terms of enrollment during 2003–04 academic year
  - Degree program
  - Class level
  - Grade point average (GPA)
  - Major field of study
- Tuition
  - Amount of tuition and fees charged
  - State prepaid or savings plan information

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Training

The training for RTI field-CADE staff was held in two separate sessions to allow for efficient use of the field staff immediately following training. Prior to these separate sessions, field supervisors participated in a telephone conference training. The field supervisors were

trained as data collectors and all participated as data collectors for the field test in 2003. The majority had prior experience as supervisors in NPSAS:2000 and were familiar with the study protocols and history. The training focused mainly on administrative responsibilities and identifying appropriate staff.

The initial field data collectors training was conducted for staff in the eastern states and Puerto Rico. The second training session was for data collectors in western states. The field supervisor training included a half-day session dealing with the project's hiring objectives and time frame, as well as supervisory and administrative responsibilities, procedures for recruiting field data collectors, and use of the systems (Case Management, Assignment and Transfer [WebATS], and e-mail). The field data collector training consisted of NPSAS:04 study objectives and time frame, an explanation of how the financial aid process works on campuses, procedures for working with the institutional coordinator and other staff at the institutions, and instruction in and practice with locating records (including review of ISIRs). The training also covered a review of and practice with each section of the CADE instrument and electronic transmission of completed cases. Finally, procedures for contacting field supervisors and other administrative procedures were discussed.

During this training, considerable use was made of location and abstraction of records using mock student case studies developed, with the assistance of National Association of Student Financial Aid Administrators (NASFAA) staff, to represent diversity in record keeping at different types of postsecondary institutions. Laptop computers were provided to all trainees for their use during training and subsequent field work. The tables of contents for the training guides used, as well as the field data collector training agenda, are included in appendix F.

All institutional coordinators, regardless of mode of CADE completion chosen, were provided with materials to assist them with CADE. A packet was sent to all institutional coordinators once the sample had been selected and CADE preloads were available that included

- a letter containing the username and password for access to the web-CADE system;
- the *NSoFaS:04, National Postsecondary Student Aid Study: NPSAS webCADE: User's Guide*, which included complete specifications, instructions, and system requirements needed for webCADE submission. Also included was a link to the institution website as well as information on alternative methods of data submission. The user's guide also discussed the study's confidentiality procedures; and
- a hardcopy list of the sampled students.

The CADE website allowed institutions to access an electronic list of the sample, which enabled them to create programs to provide the requested data from their systems for only the sampled students. All this could be done in preparation for the data entry, regardless of whether institutional staff or field staff were entering data into CADE. Several features were available from within the system to assist data entry for institutions doing self-CADE, including: help screens embedded within the program, a help desk telephone number, and an e-mail generator for problem reports. The help desk provided assistance to institutions if questions or problems arose during data entry. The help desk also provided support to institutions using the data-CADE option which generated a set of problem reports upon uploading a data file, including completed CADE information for students sampled at the institution. These reports provided comments on

any errors found in the file. The help desk ensured that institutional staff and project staff worked together to correct data while it was still being provided.

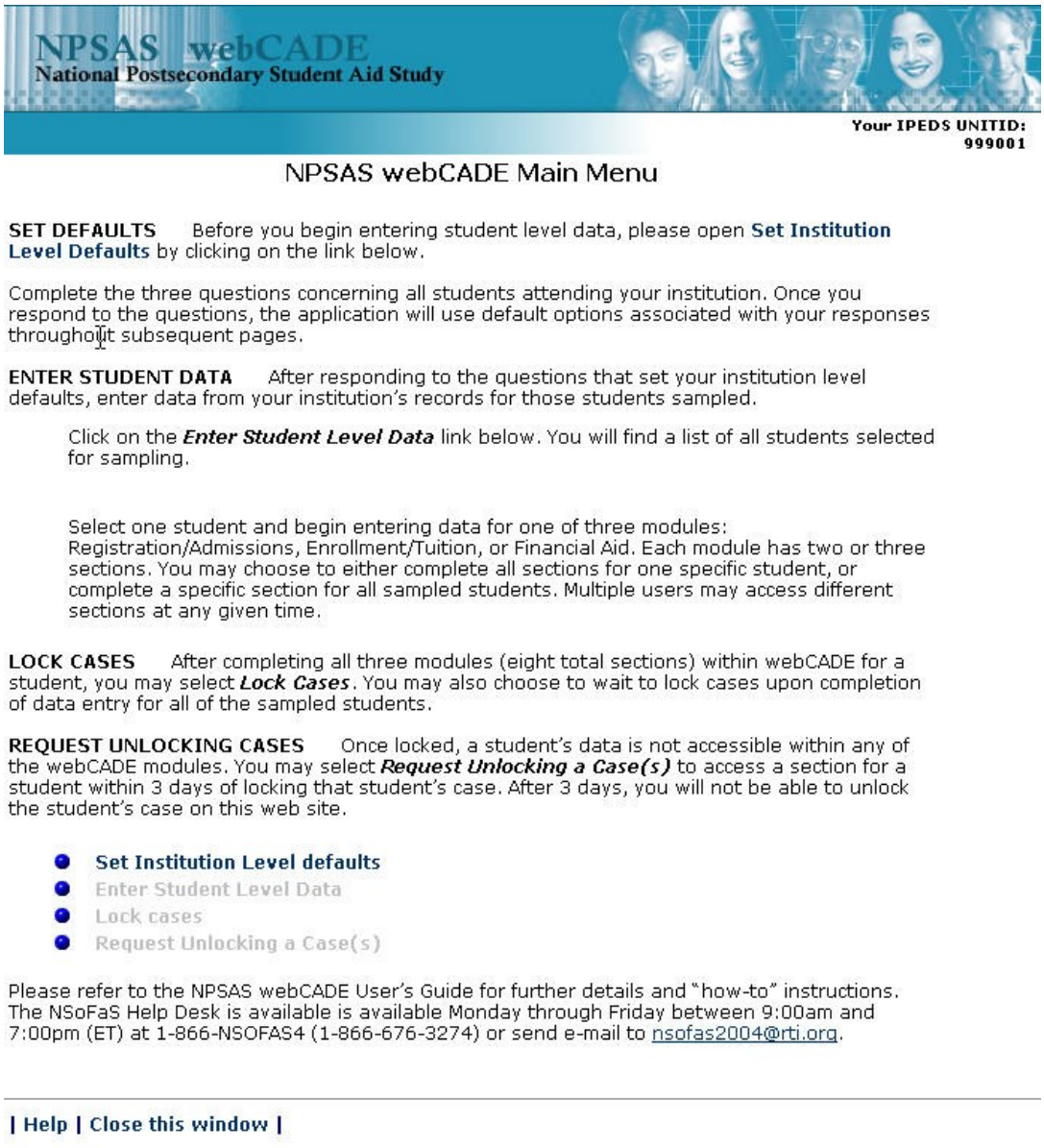
Data collection

Institutional record data for sampled students were collected using procedures similar to those successfully tested and implemented during NPSAS:2000 and during the NPSAS:04 field test. As discussed above, a web-based CADE software system was developed for use in collecting data from student records and the same CADE system was loaded onto laptops used by the RTI field data collectors for field-CADE. Institutions could choose either to enter the data themselves (self-CADE) or have an RTI-employed field data collector enter the data (field-CADE). In addition, a third option was made available for schools with programming capabilities in which electronic files could be submitted via a secured website (data-CADE). These are described in more detail below.

Self-CADE. Figure 5 presents the home page of the NPSAS CADE website. As can be seen, visitors to the website were first asked to complete their institution-level defaults (credit versus clock hour programs, grade-point average (GPA) scale, and institutional grants and scholarships). After completing these defaults, which are used by the CADE application, the user would enter all of the data for each student by clicking on the *Enter Student Level Data* link. Finally, the user would lock each case that was complete to indicate it was ready for processing. If cases were locked in error, there was a mechanism to request that a case be unlocked, provided that case had not been locked for longer than 3 days (after 3 days the user would have to call the help desk for any data changes). The website also provided the help desk phone number and e-mail address.

The home page, and all further-nested pages within the CADE application, were protected via a Secure Sockets Layer (SSL) encryption safeguard. Further security was provided by an automatic "time out" feature, through which the user was automatically logged out of the CADE application if the system was idle for 20 minutes or longer. The system did not use any persistent "cookies" (i.e., those that remain on the hard drive after the browser has been closed), thus adhering to the U.S. Department of Education's privacy policy. Selected CPS data were preloaded before data collection began to reduce data entry burden for institution staff.

Figure 5. The 2004 National Postsecondary Student Aid Study (NPSAS:04) web-CADE home page



Source: U.S. Department of Education, National Center for Education Statistics, 2004 National Study of Faculty and Students (NSoFaS:04) website.

Data-CADE. As an alternative to keying data into the web-CADE application, institutions, particularly those with large sample sizes, were given the option of submitting data files containing student record data. Explicit instructions for uploading comma-separated or delimited flat files were provided to institutions choosing this option (see appendix G). This method of data abstraction was first used in NPSAS:2000. The file specifications were customized for each institution so that they would have their own coding schemes for reporting various types of state aid and institution aid (the names of which were obtained from the institutional coordinator during the institution contacting phase of the study). Eight data files,

including student-level, term-level, and aid award-level files, were required from each data-CADE institution to accurately match the identical data structure of the database underlying the web-CADE application. Upon completion of the data-CADE file preparation, institutions submitted their data files back to RTI via the NSoFaS website. Upon submission, an automated quality control system processed the files and instantly reported back to the institutions any anomalies in the data (e.g., incorrect student ID variables, lack of term-level data for sample students, incorrect file names, etc.).

Field-CADE. Consistent with procedures implemented in past NPSAS studies, institutions were given the option of having an RTI-employed field data collector visit the institution and provide student record data-entry services at no expense to the institution. This CADE abstraction method is referred to as field-CADE.

Field data collectors used laptops with a local version of web-CADE loaded for entering data abstracted from student records. All features in the Web version were present in the laptop version, including real-time edit features to help detect out-of-range or inconsistent entries. In addition, data previously obtained from CPS were preloaded into the system before data collection began, to reduce the data collectors' level of effort. Upon completing data entry, the field data collectors transmitted the data to the same database used by web-CADE, keeping all of the completed student records together in one location.

Preloading CPS data into CADE. The first step of the CADE record abstraction process involved sending the student sample to the CPS to obtain financial aid application data. Upon completion of the CPS matching (typically a 24-hour turnaround), a number of data elements were preloaded into the CADE database, thus initializing the CADE system for that institution. These preloaded elements included an indicator of whether the student had been matched successfully to the CPS system, as well as selected CPS variables for use in CADE software edit checks. In addition, the system was customized for each institution by preloading the names of institutional financial aid programs and up to 12 state financial aid programs to assist in identifying common types of financial aid received by students.

Once CADE was initialized for a particular institution, an informational packet was sent to the designated institutional coordinator. These packets contained a listing of the students sampled and instructions for accessing the website. RTI's call center staff made follow-up phone calls to notify institutions that the CADE data collection could begin. Coordinators who previously indicated a willingness to complete the data collection via self-CADE were provided with a username and password to gain access to the web-CADE systems. As a security measure, only the coordinator was provided this password via an automatic e-mail. Based on daily status reports summarizing the progress of the self-CADE institutions, calls were made periodically to the coordinators to prompt completion of the record abstraction. Institutions using the field-CADE option were also notified by mail and contacted by the field data collector at which time an appointment was made to visit the institution.

2.3.5 Student Interview

Instrument development

The overall content of the NPSAS:04 student interview was based on items used successfully in NPSAS:2000 and NPSAS:96 in order to provide data users with the ability to make comparisons over time. Items relevant to the BPS were drawn from NPSAS:96, the last

NPSAS that served as the base year for a BPS cohort. NPSAS:2000 items specific to the B&B cohort were deleted. The NPSAS:04 instrument content was also modified to reflect changes in policy issues and topics relevant to researchers.

The student interview was developed as a web-based application, consisting of six sections grouped by topic. Figure 6 displays the structure and flow of the student instrument. The first section determined student eligibility for the NPSAS:04 study and obtained information about degree program, field of study, and enrollment history. The second section contained questions relating to student expenses and financial aid. Included in this section were items regarding employment at the NPSAS institution, such as work-study, assistantships, and fellowships. Section three focused on other employment and finances. Educational experiences such as courses taken and admission test scores were included in the fourth section, as well as items specific to BPS respondents such as first-year experiences. The fifth section of the interview gathered background and demographic information about students and their family members. The final section, applicable only to BPS respondents, requested contacting information in order to make subsequent follow-up contact in future studies.

In past rounds of NPSAS, data collection was administered by trained interviewers (primarily computer-assisted telephone interview [CATI], with some in-person interviews, or computer-assisted personal interview [CAPI]). For the first time, NPSAS:04 also included an option for self-administration via the Web. Regardless of completion mode, a single web-based instrument was employed. Mixed-mode surveys introduce benefits and challenges not experienced with single-mode surveys. Self-administration provides sample members with the ability to complete the survey at their convenience. However, interviewers are able to clarify question intent and probe when responses are unclear. Self-administered surveys require modifications to account for the mixed-mode presentation (i.e., self-administered and CATI) to maintain data quality and to make the interview process as efficient as possible for respondents. The NPSAS student interview included the following features to accommodate the mixed-mode nature of the survey:

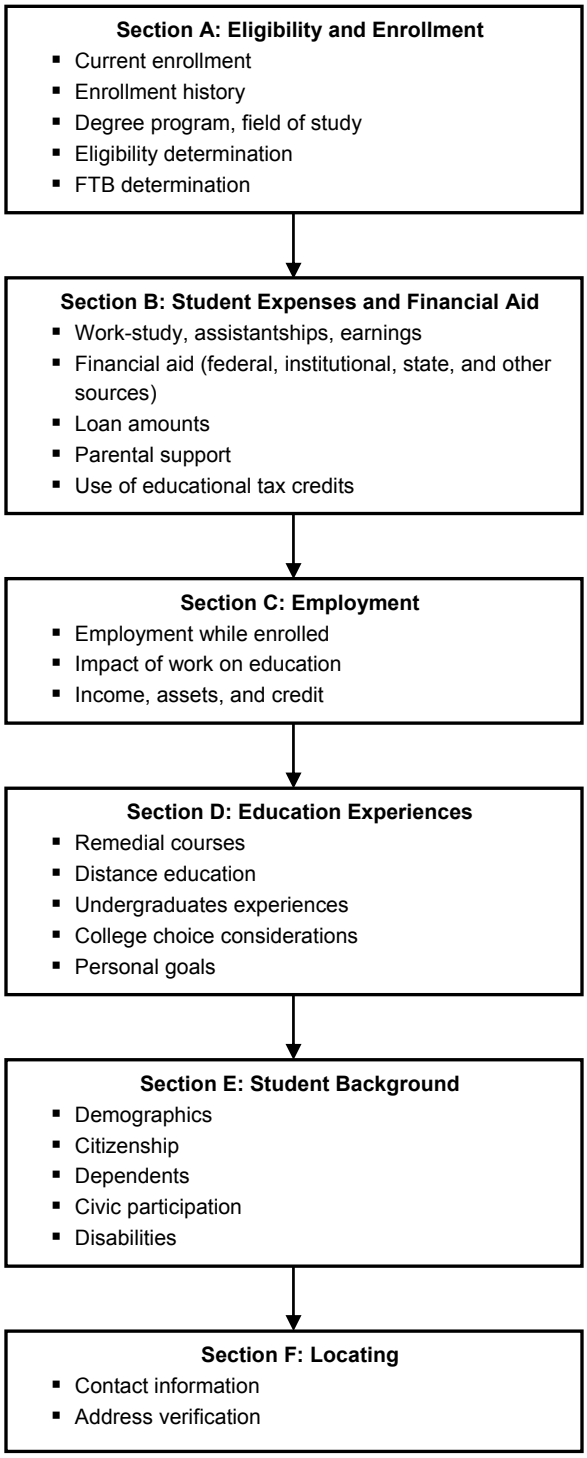
- Question wording was written so that it could be read by a respondent or read to a respondent by a telephone interviewer, while also maintaining question integrity.
- Help text was provided on all screens to assist both self-administered respondents and telephone interviewers in completing the interview.
- Pop-up boxes were displayed when out-of-range values were entered as a value for an item.
- Explicit “don’t know” responses were allowed only for items in which that was a legitimate response (such as parents’ income, use of educational tax credits, etc.). For the remaining items, respondents who did not know the answer or wished not to provide an answer could simply leave the screen blank and proceed with the interview.
- After three consecutive screens with no response, pop-up boxes were displayed to encourage participation. The prompt box reiterated the importance of the study and completeness of data, reminded sample members of the confidentiality of their responses, and requested that the respondent complete the items left blank.

With an instrument as large and complex as the NPSAS student interview, another critical factor was the determination of skip logic. Not only was it important to determine the appropriate routing from item to item on the basis of respondent status (e.g., FTB, undergraduate, graduate student), but it was also necessary to ensure that the skip logic was as efficient as possible. Sending respondents from one screen to another can add considerable transit time to web-based instruments. This increases the burden on the respondent and can lead to increased data collection costs as interviewers wait for screens to load during the interview. Another important consideration in developing the NPSAS:04 interview was the introduction of variation in response time. Web users connect through a variety of sources (e.g., dial-up, T1, high-speed cable access), use different operating systems, and have different computer resources. All of these factors were relevant to designing the instrument in order to ensure minimal burden on the respondent.

Once the instrument was programmed, rigorous testing was conducted over several iterations. Project staff and NCES staff tested numerous scenarios to evaluate the skip logic, question wording, screen layout, and efficiency of the instrument for the various student profiles expected to occur in the sample. Testing was done from a variety of locations, using a range of internet connections, and at varied times of the day to ensure that data collection would run smoothly. This process was facilitated by the use of RTI's Instrument Development and Documentation System (IDADS), which is described in detail in section 2.4.1. IDADS allowed project staff and NCES to coordinate testing efforts and provided a historical account of all problems and the solutions implemented.

An abbreviated interview was developed that contained a subset of key items from the main interview. This version was used during refusal conversion toward the end of data collection. A facsimile is presented in appendix E. The abbreviated interview was also translated into Spanish so that bilingual telephone interviewers could conduct hardcopy interviews with Spanish-speaking respondents.

Figure 6. Structure and flow of student interview: 2004



NOTE: FTB = First-time beginner.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Staff training

Various types of data collection staff were used for the NPSAS:04 student data collection, including tracing specialists, supervisors and monitors, help desk agents, and telephone interviewers. Specialized training sessions were conducted for each of these groups. A sample training agenda and table of contents from a training manual are provided in appendix F. Each training session covered an overview of the study, review of confidentiality requirements, a demonstration interview, question-by-question review of the instrument, as well as hands-on practice with the tracing module, instrument, and coding systems. In addition, each training session contained specialized instruction for each job, as described below.

- *Tracing specialists* received instruction on project-specific tracing protocols for tracing the sample members, as well as on the most effective tracing sources.
- *Supervisors and monitors* received instruction on project specific supervision and monitoring guidelines.
- *Help desk agents* received training on answering questions about the study, as well as technical questions from sample members, and were trained to document each call made to the study hotline.
- *Telephone interviewers* received information on the content of the interview, as well as on gaining cooperation from sample members, parents, and other contacts, and techniques for refusal avoidance and addressing the concerns of reluctant participants.

At the end of the project-specific training, interviewers were evaluated and certified upon successful completion of the training session.[18] The certification process involved the successful administration of the NPSAS instrument in a paired "mock" interview with a fellow trainee (one assuming the role of the interviewer and the other the sample member, and then vice versa). Trainers monitored these sessions, noting any difficulties experienced with questionnaire administration; accuracy of data entry; and voice tone, speed, and quality. In addition to successfully administering a "mock" interview, interviewers were also required to pass an oral certification exam, which focused on addressing anticipated questions and concerns from respondents.

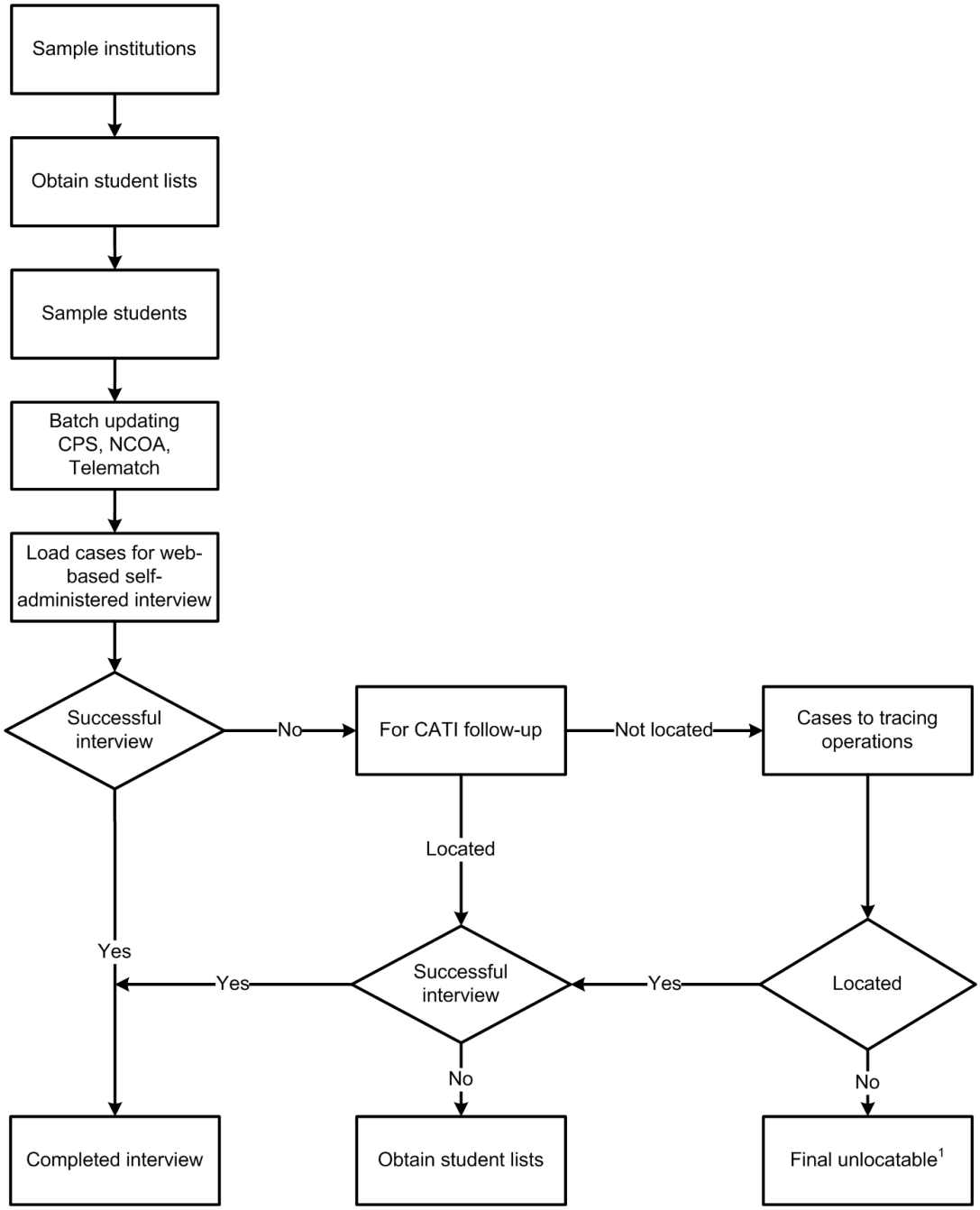
Approximately 8 weeks after the start of student interviewing, project staff and RTI Call Center Services (CCS) supervisory staff began conducting a series of refusal conversion trainings for a subset of high-performing telephone interviewers. CATI supervisors and monitors evaluated the effectiveness of telephone interviewers in dealing with respondent objections and overcoming barriers to participation. The most effective interviewers received additional and specialized instruction in specific refusal conversion techniques, including obtaining cooperation from sample members, addressing concerns raised by parents and other sample gatekeepers, validating the importance of the study, and encouraging participation among sample members who were nonrespondents prior to these conversion efforts.

[18] Certification was required of all interviewers prior to beginning work on NPSAS.

Data collection

Procedures used to locate sample members and conduct student interviews are described in the following section. Figure 7 presents the flow of activities used in locating and interviewing.

Figure 7. Overview of student data collection: 2004



[1] Even after attempts to pursue a telephone interview were exhausted, sample members could initiate and complete the student interview via the Web through the end of data collection.

NOTE: CPS = Central Processing System; NCOA = National Change of Address; CATI = Computer-assisted telephone interview.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Locating

RTI's approach to tracing sample members included two basic stages: (1) advance tracing and (2) intensive tracing. The advance tracing stage included batch database searches and lead letter mailings to sample members. The intensive tracing stage consisted of interactive tracing conducted by Call Center Services (CCS) Tracing Services. [19] The techniques described in the following sections were designed to yield the maximum number of locates with the least expense. The most cost-effective steps were taken first, minimizing the number of cases that required more costly intensive tracing efforts.

Advance tracing. Locating information obtained during institutional record abstraction was incorporated into the locator database. The data files were updated with information obtained from batch searches, from the National Change of Address (NCOA)[20] system, the Department of Education's CPS,[21] and Telematch.[22] Batch searches were conducted on a flow basis. After the locator database had been updated with the new information, a lead letter packet was mailed to the best known address for the sample member that included a standard lead letter, a study brochure, and instructions on how to access the survey via the Web (see appendix C). In the event that a sample member had moved from the mailing address in our locator database, mail forwarding from the U.S. Postal Service was requested. The most current information for the student and any other contacts were then preloaded into the CATI system.

CATI-internal locating. When assigned a case, the telephone interviewer called the telephone number designated by the system as the best number (i.e., the number among all available locator numbers that appeared to have the greatest potential for contacting the sample member) and attempted to interview the designated sample member. If the person answering the call said that the sample member could not be reached at that number, the interviewer asked the person how to contact the sample member. If this query did not provide the information needed, the interviewer initiated tracing procedures, using all information available to call other contact persons in an attempt to locate the sample member. If all tracing options available to the interviewer were exhausted without success, the case was assigned to intensive tracing via FastData,[23] or CCS Tracing Services.

[19] Tracing Services is a highly specialized unit within RTI Call Center Services (CCS) that was created in response to the recurring needs of certain research methodologies to locate large numbers of sample members. The sole focus of this unit is tracing sample members so that they can be located for research studies; the unit does not conduct any data collections.

[20] The National Change of Address (NCOA) is a database consisting of change of address data submitted to the U.S. Postal Service. Almost 100 million records are updated every 2 weeks and stored for 3 years.

[21] The Central Processing System (CPS) provides information for students who have applied for and/or received financial aid. The CPS computes student aid applicants' eligibility for student aid to assist them in attending postsecondary schools. CPS receives data from the Multiple Data Entry (MDE) contractor and sends a Student Aid Report (SAR) to the aid applicant.

[22] Telematch is a computerized residential telephone number look-up service consisting of over 65 million listings, over one million not-yet-published numbers of new movers, and over 10 million businesses. Telematch uses a name, street address, and ZIP code as search criteria and Reverse Telematch uses telephone numbers as the search criteria to provide the names under which telephones are listed.

[23] FastData is a series of database searches used to locate sample members after pre-CATI batch database searches have been done but before sending cases for intensive interactive tracing.

Intensive tracing. All cases that were not located during the advance tracing process were submitted to CCS Tracing Services for intensive locating. CCS implemented a two-tiered intensive tracing plan. The first tier identified sample members with SSNs and processed them through the following electronic databases.[24]

- *Query of Credit Bureau databases*. **Equifax**, a credit bureau that maintains credit files on a large number of individuals; **Experian,** which holds more demographic and credit information on individuals and businesses than any other company in the world and **TransUnion,** which also holds demographic and credit information on individuals and businesses, were all used to locate sample members.
- *Query of internet databases*. Contractor staff had direct electronic access to various databases, which included names, SSNs, and current and former addresses and telephone numbers of individuals.
- *Query of the Select Phone Book CD-ROM data*. This database contains every published telephone number in the United States, with associated names and addresses. It can be sorted within city by address, to obtain telephone numbers and names of neighbors.

New telephone numbers generated from the above searches were sent back into the Case Management System for telephone interviewing. If a new address was generated, but no telephone number, tracers used directory assistance or other databases to obtain telephone numbers. This first level of effort minimized the time that cases were out of production.

All remaining cases (those lacking new information from the SSN search) underwent a more intensive level of tracing in the second-tier approach. This approach involved the following procedures: (1) checking directory assistance for telephone listings at various addresses; (2) using electronic reverse-match databases to obtain the names and telephone numbers of neighbors and then calling the neighbors; (3) calling persons with the same unusual surname in small towns or rural areas to see if they were related to or knew the sample member; (4) contacting the current or last-known residential sources such as neighbors, landlords, current residents, tax assessors, realtors, and other business establishments related to previous addresses associated with the sample member; (5) calling colleges, military establishments, and correctional facilities to follow up on leads generated from other sources; and (6) checking various tracing websites. Tracers checked new leads produced by these tracing steps to confirm the address and telephone numbers for the sample members. When the information was confirmed, the case was returned to the CMS for completion. If the information could not be confirmed (e.g., there were no working telephone numbers or numbers for relevant neighborhood sources were unpublished), no further attempts were made to locate such sample members.[25]

Notification materials and student resources

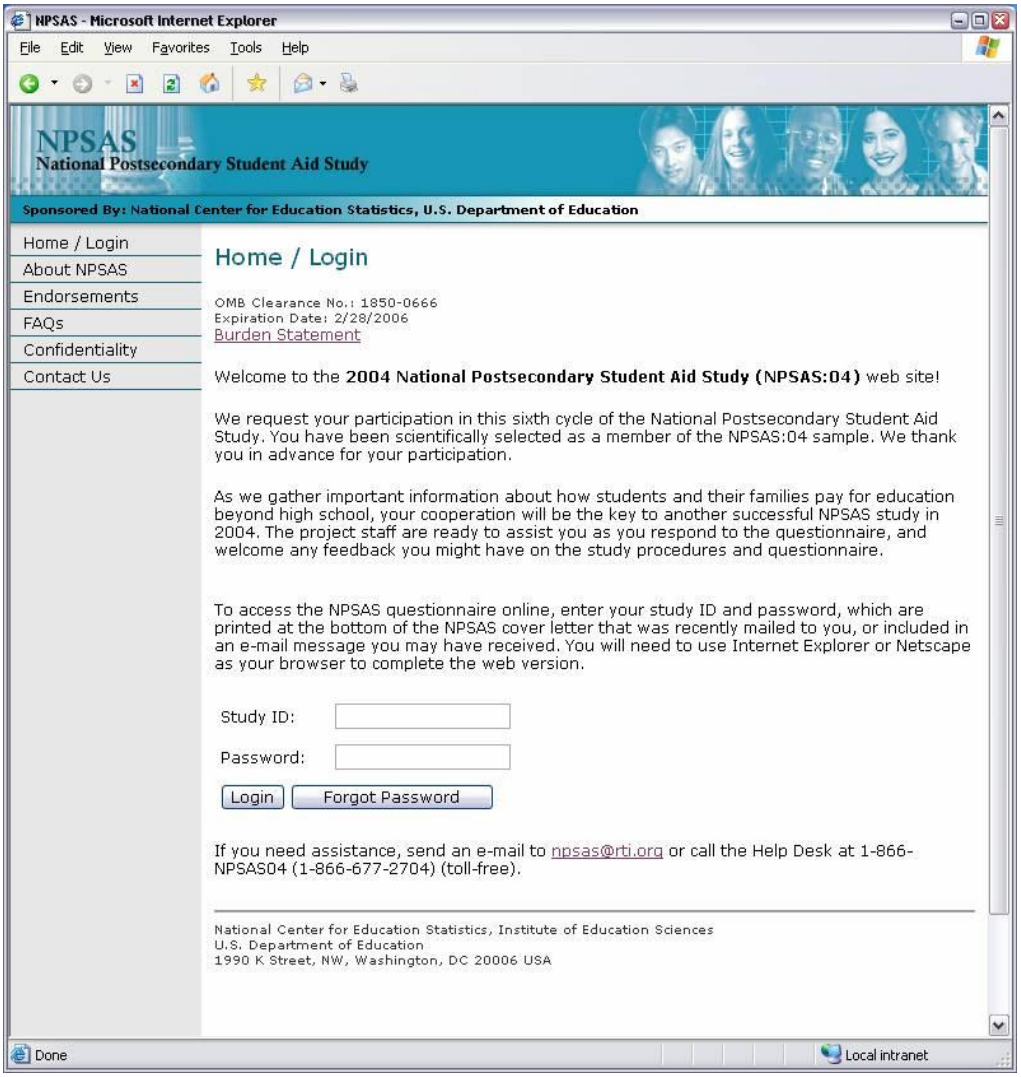
Student website. A study website was designed for students. The website provided important information about NPSAS:04, such as the purpose and history of the study and a summary of findings from prior interviews. Confidentiality procedures were described and the

[24] Tracing activities were restricted to the collection of locating/directory information.
[25] Unlocatable sample members always had the opportunity to complete self-administered interview or to call in for a telephone interview through the end of data collection.

use of the data was explained. It also provided contact information for the study's help desk and project staff and links to the NCES and RTI websites. The website also provided a link from which sample members could log in to the student interview. The website address was included with all mailings to sample members.

The NPSAS:04 website (figure 8) was designed in accordance with NCES Web policies. A two-tier security approach was used to protect all address and interview data collected through the website. At the first tier, sample members were required to log on to the secure areas of the website using a unique and randomly assigned study ID and password sent by mail. At the second tier of security, data entered on the NPSAS:04 website—both contact information and interview responses—were protected with SSL technology, ensuring that only encrypted data were transmitted over the Internet. As an additional security measure, the interview contained an automatic "time-out" feature through which a respondent was logged-out if the system was idle for 30 minutes.

Figure 8. The 2004 National Postsecondary Student Aid Study (NPSAS:04) website home page



Source: U.S. Department of Education, National Center for Education Statistics, 2004 National Study of Faculty and Students (NSoFaS:04) website.

Help desk. The help desk staff was available to assist sample members who had questions or problems accessing and/or completing the self-administered interview. A toll-free hotline was set up to accept incoming help desk calls. If technical difficulties prevented a sample member from completing a self-administered interview, a help desk staff member, who was also trained to conduct telephone interviews, would encourage him/her to complete a telephone interview rather than to attempt the self-administered interview.

The help desk application documented all incoming calls from sample members. In addition to this primary documentation function, it provided the following:

- information needed to verify a sample member's identity to assist with login difficulties;
- login information allowing a sample member to access the Web interview; and
- means for tracking problems that could not be immediately resolved.

The help desk application also provided project staff with various reports on the type and frequency of problems experienced by sample members, as well as a way to monitor the resolution status of all help desk inquiries.

Lead letter mailing. Once a valid address for a sampled student was identified either through the participating institution or a batch database search, each sample member was mailed a lead letter. The personalized lead letter signed by the NCES commissioner provided information about the study, a description of the options for completing the questionnaire via the Web or telephone, the electronic address (URL) for the project website, and the sample member's username and password for secure access to the website. A study brochure was also included with the mailing.

The letter was used to inform sample members that they were eligible to complete the NPSAS:04 interview at their convenience on the Web and provided them with the technical information on how to do so. The letter also provided an e-mail address and the NPSAS:04 toll-free telephone number to the help desk as a means for sample members to update their contact information, schedule an appointment, or complete the interview by telephone. Lead letter mailouts began in early February 2004 and by the end of July 2004, 138,320 lead letter packets had been mailed.

Electronic mail (E-mail). E-mail was an important tool in the locating and interviewing process. In addition to sending a lead letter mailing, students were sent a lead e-mail as an additional way of making initial contact. The content of this e-mail mirrored the content of the lead letter but also included a hyperlink Web address so students were able to click on the address to be taken directly to the Web survey. E-mail follow-up messages were sent to sample members with valid e-mail addresses 1 day, 7 days, and 14 days after the initial hard-copy mailing.

E-mail was also used as a tool for locating hard-to-reach sample members and for prompting participation among nonrespondents. The e-mail messages were used to encourage sample members either to complete the web-based self-administered survey or to contact RTI to complete the survey or to set an appointment for a telephone interview.

Student interviews

Self-administered interviews. The data collection notification materials invited sample members to log into the study website and provided all the information needed to do so. During the 4 weeks immediately following the notification letter, only self-administered interviews via the Web were completed unless a student called in to the help desk for assistance and completed the telephone interview. Outbound calls by interviewing staff were not initiated until sample members had sufficient opportunity to complete the interview. E-mail prompts were sent to sample members periodically during the 4-week period to encourage participation and remind them of the address for the study website, as well as the toll-free help desk telephone number.

Sample members were assured of the confidentiality of their responses. They were also informed of the voluntary nature of the survey, noting that they could decline to answer any survey question. Furthermore, the convenience features of the web-based survey were emphasized—especially that the survey could be completed at any time from any location with internet access and that respondents could break off and resume the interview if needed. The web interview site remained available 24 hours per day, 7 days per week throughout the entire data collection period. This availability gave sample members the option to complete interviews online during the entire data collection period.

Telephone interviews. Attempts to locate and interview study sample members who had not yet completed an interview began 4 weeks after sample members were invited to complete the self-administered interview. Once located, an attempt was made to conduct the full interview with the sample member. However, some cases required special treatment. To deal with those who initially refused to participate (including locator sources who acted as "gatekeepers," preventing access to the sample member), certain interviewers were trained in refusal conversion techniques. Sample members and their locator sources who spoke only Spanish, primarily located in Puerto Rico, were assigned to bilingual CATI interviewers.

Use of incentives. In an effort to increase study response rates, sample members were offered an incentive of $10 for completing a self-administered interview in the first 4 weeks. Sample members would receive the incentive regardless of participation mode, provided that they completed their interview prior to their individual deadline.

Toward the end of data collection, a different incentive plan was used with particular types of nonrespondents: (1) cases where the sample member initially refused the interview; (2) sample members for whom intensive tracing yielded a good mailing address, but no telephone number; and (3) cases identified as "hard to reach" (i.e., those with 20 or more call attempts, where contact had been established with the sample member and no "hard" appointment was pending). The incentive offer consisted of a letter from the project director on RTI letterhead, or an e-mail tailored to the specific type of nonrespondent (i.e., refusal or hard to reach/no telephone number). Respondents were promised a check for $20 if they completed the interview, regardless of the mode they used to do so. The incentive letters were mailed on a flow basis as respondents met one of the three criteria described above.

Finally, in an effort to convert sample members who still had not responded to the previous incentive offers near the end of data collection, all pending cases received a final letter requesting participation by the end of the data collection period. Postcards and letters reemphasized the importance of the study and offered all remaining respondents a check for $30 if they completed the interview, regardless of the mode they used to do so.

2.3.6 Data Quality Evaluation

All stages and components of NPSAS:04 were carefully monitored and evaluated throughout the course of development and production. Table 7 outlines some of the major evaluations conducted as part of the full-scale study.

Table 7. Summary of NPSAS:04 evaluations

Major area of evaluation	Evaluation approaches
Training for data collection	Debrief field abstractors.[1]
	Debrief computer-assisted telephone interview (CATI) staff.[1]
Enrollment list acquisition	Analyze overall response rate, accuracy, and time to produce lists.
Student record abstraction	Analyze overall outcomes, including institutional participation, nonresponse, and refusal
	Analyze data quality (missing data) under conditions of web-CADE, field-CADE, and data file production approaches.
	Debrief institutional coordinators.[1]
	Debrief field staff. [1]
Student tracing and locating activities	Debrief tracing staff and supervisors.[1]
	Analyze all sources and levels of tracing results and costs.
Student interviewing	Analyze quality control monitoring data.
	Analyze CATI operational parameters (e.g., numbers of calls per case, total interviewer hours per completed interview).
	Analyze interview response burden, overall and by section.
	Debrief interviewers, monitors, and supervisors.[1]
	Analyze response rates and patterns of interview nonresponse, overall and by mode of administration.
	Analyze impact of financial incentive on response rate.
Nonresponse bias analysis	Analyze nonresponse bias at the following levels: institutional, student, and item.

[1] Informal debriefings of staff involved in different data collection tasks were conducted throughout the study. Information gathered through these debriefings was used to enhance understanding of the outcomes of more formal evaluations and is therefore not described separately in this report.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

2.4 Data Collection Systems

2.4.1 Instrument Development and Documentation System (IDADS)

IDADS is a controlled web environment in which project staff developed, reviewed, modified, and communicated changes to specifications, code, and documentation for the NPSAS:04 student interview. All information relating to the instrument was stored in an SQL server database and was made accessible through Windows™ and Web interfaces. IDADS contains three modules: specification, programming, and documentation.

Initial specifications were generated within the IDADS *specification module*. This module enabled access for searching, reviewing, commenting on, updating, exporting, and importing information associated with instrument development. All records were maintained individually for each item, which provided a historical account of all changes requested by both project staff and NCES.

Once specifications were finalized, the *programming module* within IDADS produced hypertext transfer markup language (HTML), Active Server Pages (ASPs), and JavaScript template program code for each screen based on the contents of the SQL Server database. This output included screen wording, response options, and code to write the responses to a database, as well as code to automatically handle such web-instrument functions as backing up and moving forward, recording timer data, and linking to context-specific help text. Programming staff edited the code that was automatically generated by this module to customize screen appearance and program response-based routing.

The *documentation module* contained the finalized version of all instrument items, the screen wording for each, and variable and value labels. Also included in this module were the more technical descriptions of items such as variable types (alpha or numeric), information regarding to whom the item was administered and to whom the item applied, and frequency distributions for response categories. The documentation module was used to generate the student interview facsimile and the associated documentation files to be used as input to the VTS (discussed in section 2.4.3).

2.4.2 Integrated Management System (IMS)

The IMS is a comprehensive set of desktop tools designed to give project staff and NCES easy access to a centralized repository for project data and documents. The NPSAS:04 IMS was developed based on a framework initially developed (and refined) under previous NCES studies conducted by RTI. These include NPSAS:2000, B&B:2000/01, and B&B:93/03. As with these previous studies, the NPSAS:04 IMS consisted of independent, but integrated, modules. To the extent possible, the NPSAS:04 IMS was developed using commercial, nonproprietary PC-based software systems.

The major modules of the NPSAS:04 IMS include the following:

IMS website

- Contains tools and strategies to assist project staff and the NCES project officer in managing the study. All information pertinent to the study is located there, accessible via the Web, in a secure desktop environment. Available on the IMS are the current project schedule, monthly progress reports, daily data collection reports and status reports, project plans and specifications, key project information and deliverables, instrument specifications, staff contacts, the project bibliography, and a document archive. The IMS also has a download area from which the client and subcontractors can retrieve files when necessary.
- Infrastructure was programmed in ASP.
- SQL Server 2000 serves as the back-end database where applicable (maintaining the project staff contact list, Technical Review Panel membership, confidentiality reports, etc.)

Receipt Control System (RCS)

- An integrated set of systems that monitors all activities related to data collection, including tracing and locating. Through the RCS, project staff are able to perform stage-specific activities, track case statuses, identify problems early, and implement solutions effectively. RCS locator data were used for a number of daily tasks related to sample maintenance.
- Back-end database is Microsoft SQL Server 2000.
- Front-end interface and reports were programmed in ASP and SQL Server Reports Server.

Institution Contacting System (ICS)

- The ICS allows staff to log all contacts with institutions and determine the next steps for staff working with specific institutions. From within the ICS, the mailout program produces mailings to sample members, the electronic mailout program produces e-mail notifications and reminders to sample members, the tracing program enables staff to send and receive tracing information from locating firms, the query system enables administrators to review the locator information and status for a particular case, and the mail return system enables project staff to update the locator database. The RCS also interacts with the Case Management System (discussed below) and the CCS Tracing Services databases, sending locator data between the three systems as necessary.
- Back-end database is Microsoft SQL Server 2000.
- Front-end interface and reports were programmed in ASP and SQL Server Reports Server.

Case Management System (CMS)

- The technological infrastructure that connects the various components of the CATI system, including the student questionnaire, utility screens, databases, call scheduler, report modules, links to outside systems, and other system components. It utilizes a call scheduler to assign cases to interviewers in a predefined priority order. In addition to delivering appointments to interviewers at the appropriate time, the call scheduler also calculates the priority scores (the order in which cases need to be called based on preprogrammed rules), sorts cases in nonappointment queues, and computes time zone adjustments to ensure that cases are not delivered outside the specified calling hours. The call scheduler also permits callbacks to be set, and assigns status codes to the case. In addition, each case contains one or more roster lines that detail specific contact information for a case (e.g., home phone number, work phone number, etc.). The call scheduler uses a call algorithm based on the previous call results to determine which roster line should be called next.
- Back-end database is SQL Server 2000.
- Infrastructure was programmed in Visual Basic (VB).

Student instrument

- Back-end database is SQL Server 2000.
- Instrument was programmed in ASP.
- Edit checks were programmed using JavaScript.
- Web security was implemented using SSL certification with 128-bit encryption.
- Users' browsers were required to support, and be enabled for, JavaScript and session cookies (i.e., those that are erased from the hard drive after the browser has been closed).
- Final student interview database was maintained in SAS 8 (subsequently upgraded to SAS 9.1).
- Student status and summary reports were programmed in SAS 8 (subsequently upgraded to SAS 9.1).

CADE

- Back-end database is Microsoft SQL Server 2000.
- Front-end interface was programmed in ASP.
- Edit checks were programmed using JavaScript.
- Reports were developed using ASP.
- Web security was implemented using SSL certification with 128-bit encryption.
- Users' browsers were required to support, and be enabled for, JavaScript and session cookies (i.e., those that are erased from the hard drive after the browser has been closed).
- Final CADE database was maintained in SAS 8 (subsequently upgraded to SAS 9.1).

Automated processing

During data collection, a series of automated batch files were executed nightly via Windows XP scheduled processing to ensure that project staff were able to closely monitor progress during all stages of data collection. These automated processes included the following:

- **Dataload.** This program contained many different subprocesses, with the overall purpose being to process transactions generated during the day by various project systems and activities, and post the transactions to the RCS, updating institution and student-level case status information. Transactions included results from enrollment list processing, sampling, CPS matching, CADE preload and data receipt processing, lead-letter mailout and return, and student instrument preloading and interviewing.
- **RCS report generator.** Each night following the completion of the dataload process, the RCS report generator created HTML pages detailing both the institution- and student-level current status reports. It also produced miscellaneous project management reports including: Abstraction Method Report, Enrollment (list type) Report, Chief Administrator Participation Report, Enrollment List Acquisition

Report, CADE Status Summary Report (overall and for the BPS cohort), and Student Interview Summary Reports. The process automatically posted these reports to the IMS.

- **Data upload to master files.** Each night this process would update master files containing CADE and student interview data with newly acquired data, including complete and partial cases.
- **Data processing.** Separate programs ran nightly to edit the raw CADE and student interview data (see chapter 5 for more detail).

2.4.3 Variable Tracking System (VTS)

The central mechanism for constructing input files for the NCES Electronic Codebook (ECB) was a software application called the Variable Tracking System (VTS). The VTS tracked and stored documentation for both interview and derived variables required for the ECB and NCES' Data Analysis System (DAS). This included weighted and unweighted variable distributions, variable labels, value codes and labels, and a text field describing the development and source of each variable and, if applicable, the programming code used to construct it. Input files for the ECB and DAS systems were automatically produced by the VTS according to NCES specifications.

Chapter 3
Data Collection Outcomes

This chapter summarizes the results of the various stages of data collection implemented in the 2004 National Postsecondary Student Aid Study (NPSAS:04). Study response rates for institutions and students are presented first. Next, completion rates for individual data sources are discussed, including rates of matching to extant databases, locating results, and interviewing outcomes (by mode of survey administration).

3.1 Institutional Participation

Eligible sample institutions were asked to participate in two stages of NPSAS:04 by (1) providing a comprehensive list of enrolled students for sample selection and (2) providing data from student records for the sampled students. Consequently, the potential for institutional nonresponse existed at these two points in the survey process. Rates of institutional response—for the national and state-representative samples—are discussed in the following sections.

3.1.1 National Sample

Counts of eligible institutions in the national sample are shown in table 8, by institutional level, institutional control, and type of institution. About 1,630 of the 1,670 institutions initially selected for the full-scale study were determined to be eligible for NPSAS:04. Table 8 also shows that about 1,360 (84 percent) of the 1,630 eligible sample institutions provided a list of enrolled students that could be used for sample selection.[26] List provision rates (among eligible institutions) varied by type of institution, ranging from 77 percent for public less-than-2-year institutions to 89 percent for private not-for-profit less-than-4-year institutions. Weighted participation rates were calculated based on the institutional probabilities of selection and enrollment[27] and are also shown in table 8.[28] The overall weighted participation rate was 80 percent.[29]

[26] Two institutions provided a list that had no student identifying information, and the institutions were not willing to provide this information. These lists were not sufficient for sample selection.

[27] In addition to the probabilities of selection, the participation rates accounted for the institution enrollment from the 2003 Integrated Postsecondary Education Data System (IPEDS). The unit of analysis for NPSAS is a student, so factoring the enrollment into the participation rates, gives an indication of what percentage of students are represented by the participating institutions.

[28] The weighted response rates can be interpreted as the estimated percentages of institutions in the population that would have provided a usable student sampling list, if asked.

[29] When the weighted response rates differ from the unweighted response rates, it is due to the nonresponding institutions having, on average, a lower or higher weight than the responding institutions. NPSAS:04 was designed to produce efficient estimates only at the student level. Institutions were selected with probabilities proportional to size; therefore, weighted institution-level estimates are subject to a high level of sampling variation.

Table 8. Numbers of NPSAS:04 sampled, eligible, and participating institutions and enrollment list participation rates, by institutional characteristics: national sample

Institutional characteristics[2]	Sampled institutions	Eligible institutions[3]	Institutions providing lists[1]		
			Number	Unweighted percent	Weighted percent
All institutions	1,670	1,630	1,360	83.5	80.0
Institutional level					
Less-than-2-year	260	250	200	82.1	80.8
2-year	490	480	410	85.4	78.0
4-year non-doctorate-granting	460	460	380	83.3	74.6
4-year doctorate-granting	450	450	370	82.4	85.6
Institutional control					
Public	810	800	680	84.9	79.6
Private not-for-profit	570	560	450	81.2	79.8
Private for-profit	290	270	230	84.2	86.7
Type of institution					
Public less-than-2-year	70	60	50	76.6	74.3
Public 2-year	380	380	320	85.4	77.6
Public 4-year non-doctorate-granting	130	130	110	85.1	70.3
Public 4-year doctorate-granting	230	230	200	86.3	87.1
Private not-for-profit less-than-4-year	70	70	70	89.0	92.6
Private not-for-profit 4-year non-doctorate-granting	280	270	220	81.9	78.1
Private not-for-profit 4-year doctorate-granting	220	220	170	77.7	80.8
Private for-profit less-than-2-year	170	160	140	84.0	82.3
Private for-profit 2-year or more	110	110	90	84.4	88.2

[1] Percents are based on the eligible institutions within the row under consideration.

[2] Institutional characteristics are based on data from the sampling frame which was formed from the 2000–01 and 2002–03 Integrated Postsecondary Education Data System (IPEDS).

[3] Among the 30 ineligible institutions: 10 closed after the sampling frame was defined, and 10 failed to meet one or more of the criteria for institutional NPSAS eligibility. The remainder were treated as merged institutions because two or more campuses were included on one combined student list.

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

3.1.2 State Samples

Counts of eligible institutions for the state samples are shown in table 9, by state and type of institution. Table 9 also shows the weighted and unweighted enrollment list provision rates (among eligible institutions), which varied by state and type of institution considered. The weighted participation rate ranged from 53 percent to 100 percent.

Table 9. Numbers of NPSAS:04 sampled, eligible, and participating institutions and enrollment list participation rates, by institutional characteristics: state samples

Institutional characteristics[2]	Sampled institutions	Eligible institutions	Institutions providing lists[1]		
			Number	Unweighted percent	Weighted percent
California					
Public 2-year	40	40	30	65.8	66.2
Public 4-year	30	30	20	57.6	53.3
Private not-for-profit 4-year	40	40	30	73.2	77.6
Connecticut					
Public 2-year	20	20	10	86.7	99.9
Public 4-year	10	10	10	100.0	100.0
Private not-for-profit 4-year	20	20	20	85.0	99.2
Delaware					
Public 2-year	#	#	#	100.0	100.0
Public 4-year	#	#	#	100.0	100.0
Private not-for-profit 4-year	#	#	#	100.0	100.0
Georgia					
Public 2-year	30	30	30	100.0	100.0
Public 4-year	20	20	20	100.0	100.0
Private not-for-profit 4-year	30	30	30	81.8	91.1
Illinois					
Public 2-year	30	30	30	90.0	90.0
Public 4-year	10	10	10	83.3	87.7
Private not-for-profit 4-year	30	30	20	75.9	77.1
Indiana					
Public 2-year	20	20	20	100.0	100.0
Public 4-year	10	10	10	100.0	100.0
Private not-for-profit 4-year	30	30	30	90.0	92.0
Minnesota					
Public 2-year	30	30	30	100.0	100.0
Public 4-year	10	10	10	100.0	100.0
Private not-for-profit 4-year	40	40	30	85.7	94.6
Nebraska					
Public 2-year	10	10	10	71.4	67.2
Public 4-year	10	10	10	85.7	97.0
Private not-for-profit 4-year	20	20	10	86.7	68.2
New York					
Public 2-year	40	40	30	86.5	93.0
Public 4-year	30	30	30	93.1	94.2
Private not-for-profit 4-year	40	40	30	81.6	82.6
Oregon					
Public 2-year	20	20	10	82.4	85.3
Public 4-year	10	10	10	90.0	97.0
Private not-for-profit 4-year	30	20	20	91.7	94.7
Tennessee					
Public 2-year	20	20	20	100.0	100.0
Public 4-year	10	10	10	100.0	100.0
Private not-for-profit 4-year	30	30	20	80.0	84.6
Texas					
Public 2-year	30	30	30	93.1	91.5
Public 4-year	30	30	30	86.7	87.7
Private not-for-profit 4-year	30	30	20	80.0	84.4

Rounds to zero.

[1] Percents are based on the eligible institutions within the row under consideration.

[2] Institutional characteristics are based on data from the sampling frame which was formed from the 2000–01 and 2002–03 Integrated Postsecondary Education Data System (IPEDS).

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

3.2 Study Respondents

As noted in the previous chapter, student-level data for NPSAS:04 are collected from a variety of sources, including student records (computer-assisted data entry [CADE]), student interviews, and extant federal and private databases (Central Processing System [CPS], National Student Loan Data System [NSLDS], ACT, and SAT files). For NPSAS:04, a definition of the minimum data requirements, regardless of source, to be considered a study respondent was adopted. Specifically, a study respondent is defined as any sample member who is determined to be eligible for the study (based on the eligibility criteria specified in chapter 2) and, minimally, has valid data from any source for the following:

- student type (undergraduate or graduate/first professional);
- date of birth or age;
- gender; and
- at least 8 of the following 15 variables:
  - dependency status;
  - marital status;
  - any dependents;
  - income;
  - expected family contribution (EFC);
  - degree program;
  - class level;
  - first-time beginner (FTB) status;
  - months enrolled;
  - tuition;
  - received federal aid;
  - received non-federal aid;
  - student budget;
  - race; and
  - parent education.

Student-level study response rates for both the national sample and the state samples are presented below.

3.2.1 National Sample

Counts of eligible students are shown in table 10, by type of institution. About 8,200 (8 percent) of the 109,210 students initially selected for the full-scale study were determined to be ineligible for NPSAS:04. Upon the completion of data collection, 90 percent of the 101,010 eligible sample members had sufficient key data to be classified as study respondents. Weighted

response rates were calculated based on the institutional weights and student probabilities of selection and are also shown in table 10.[30] The student weighted response rate was 91 percent.[31]

Table 10 also shows that the unweighted student response rates (among eligible students) varied by type of institution, ranging from 81 percent for students from public 2-year institutions to 96 percent for students from private not-for-profit 4-year non-doctorate institutions. Response rates also varied by student type: 91 percent for FTBs, 87 percent for other undergraduates, and 94 percent for graduate and first-professional students.

Table 10. Numbers of NPSAS:04 sampled and eligible students and response rates, by institutional characteristics and student type: national sample

Institutional characteristics and student type[3]	Sampled students	Eligible students[4]	Responding students[1,2]	
			Unweighted percent	Weighted percent
All students	109,210	101,010	89.8	91.0
Institutional level				
Less-than-2-year	13,320	11,330	92.6	93.6
2-year	41,510	37,290	83.1	84.6
4-year non-doctorate-granting	21,450	20,550	94.0	94.8
4-year doctorate-granting	32,930	31,840	94.1	94.7
Institutional control				
Public	71,030	65,540	87.0	89.2
Private not-for-profit	22,730	21,660	95.3	96.1
Private for-profit	15,460	13,820	95.0	96.1
Type of institution				
Public less-than-2-year	3,180	2,580	84.2	90.6
Public 2-year	36,300	32,450	81.3	83.9
Public 4-year non-doctorate-granting	9,200	8,880	91.9	93.3
Public 4-year doctorate-granting	22,350	21,620	93.7	94.2
Private not-for-profit less-than-4-year	3,060	2,770	94.3	94.6
Private not-for-profit 4-year non-doctorate	9,740	9,300	96.3	96.9
Private not-for-profit 4-year doctorate-granting	9,930	9,590	94.5	95.4
Private for-profit less-than-2-year	9,270	8,030	94.9	94.3
Private for-profit 2-year or more	6,190	5,790	95.0	96.7
Student type				
Total undergraduates	97,090	89,480	89.3	90.3
Potential FTB	49,410	44,670	91.2	91.4
Other undergraduates	47,680	44,810	87.3	90.0
Graduate/first professional	12,120	11,530	94.2	95.1

[1] A responding student is defined as any eligible student for whom sufficient data were obtained from one or more sources, including student interview, institutional records, and the Department of Education's Central Processing System (CPS).

[2] Percents are based on the eligible students within the row under consideration.

[3] Institutional characteristics are based on data from the sampling frame which was formed from the 2000–01 and 2002–03 Integrated Postsecondary Education Data System (IPEDS). Student type is based on data from the sampling frames which were the enrollment lists received from participating institutions.

[4] Ineligible students were identified during the student interview or from institutional records if student eligibility was not determined from a student interview.

NOTE: Detail may not sum to totals because of rounding. FTB = first-time beginner.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

[30] The weighted response rates can be interpreted as the estimated percentages of students in the population that would have responded, if asked.

[31] When the weighted response rates differ from the unweighted response rates, it is due to the nonresponding students having, on average, a lower or higher weight than the responding students.

3.2.2 State Samples

Counts of eligible students for the representative undergraduate state samples are shown in table 11, by state and type of institution. Table 11 also shows that the unweighted and weighted response rates (among eligible students) varied by state and type of institution. The weighted response rates range from 61 percent to 100 percent.

Table 11. Numbers of NPSAS:04 sampled, eligible, and responding students and response rates, by institutional characteristics: state samples

Institutional characteristics[3]	Sampled undergraduate students	Eligible undergraduate students[4]	Responding undergraduate students[1,2]	
			Unweighted percent	Weighted percent
California				
Public 2-year	5,390	4,800	59.4	64.5
Public 4-year	1,390	1,400	90.2	91.3
Private not-for-profit 4-year	840	820	90.1	91.1
Connecticut				
Public 2-year	610	570	73.9	77.5
Public 4-year	520	500	99.0	99.1
Private not-for-profit 4-year	490	480	95.4	95.5
Delaware				
Public 2-year	900	840	55.0	60.6
Public 4-year	610	590	91.6	98.3
Private not-for-profit 4-year	540	530	85.7	93.0
Georgia				
Public 2-year	2,110	1,930	90.5	93.4
Public 4-year	940	910	95.4	94.9
Private not-for-profit 4-year	470	450	98.2	98.2
Illinois				
Public 2-year	1,950	1,640	86.4	88.8
Public 4-year	740	730	95.2	95.5
Private not-for-profit 4-year	740	710	95.9	96.5
Indiana				
Public 2-year	410	350	98.3	98.4
Public 4-year	930	920	92.8	95.1
Private not-for-profit 4-year	560	540	98.3	98.5
Minnesota				
Public 2-year	1,430	1,320	60.0	61.0
Public 4-year	630	620	79.5	78.4
Private not-for-profit 4-year	510	500	97.8	98.8
Nebraska				
Public 2-year	450	390	97.0	98.2
Public 4-year	520	510	95.7	94.4
Private not-for-profit 4-year	340	330	100.0	100.0
New York				
Public 2-year	2,200	2,030	83.6	87.1
Public 4-year	1,490	1,450	85.2	84.7
Private not-for-profit 4-year	1,660	1,590	96.6	96.8
Oregon				
Public 2-year	1,040	920	90.7	92.8
Public 4-year	640	630	97.1	97.7
Private not-for-profit 4-year	340	330	98.2	99.4

See notes at end of table.

Table 11. Numbers of NPSAS:04 sampled, eligible, and responding students and response rates, by institutional characteristics: state samples—Continued

Institutional characteristics[3]	Sampled undergraduate students	Eligible undergraduate students[4]	Responding undergraduate students[1,2]	
			Unweighted percent	Weighted percent
Tennessee				
Public 2-year	1,120	960	89.2	90.5
Public 4-year	790	760	90.9	93.3
Private not-for-profit 4-year	390	390	96.4	95.3
Texas				
Public 2-year	3,170	2,800	88.2	90.8
Public 4-year	1,620	1,560	95.7	96.0
Private not-for-profit 4-year	640	620	98.6	98.4

[1] A responding student is defined as any eligible student for whom sufficient data were obtained from one or more sources, including: student interview, institutional records, and the Department of Education's Central Processing System (CPS).

[2] Percents are based on the eligible students within the row under consideration.

[3] Institutional characteristics are based on data from the sampling frame which was formed from the 2000–01 and 2002–03 Integrated Postsecondary Education Data System (IPEDS).

[4] Ineligible students were identified during the student interview or from institutional records if student eligibility was not determined from a student interview.

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

3.3 Data Collection Results, by Source

Chapter 2 described the various sources of data for NPSAS:04 and the methods through which they were obtained. The following section presents the results of each stage of data collection. This section presents results for individual data sources. An individual sample member's status as a study respondent was determined by the amount of data across sources (see Section 3.2 for a definition of the requirements for study respondent classification). Thus, rates presented for the following data sources do not correspond to study response rates.

3.3.1 Student Record Matching

Central Processing System (CPS)

Table 12 summarizes the results of matching and downloading student data from the U.S. Department of Education's CPS. The CPS contains data provided to the U.S. Department of Education by students and their families when they complete the Free Application for Federal Student Aid (FAFSA). Therefore, successful matching to CPS can only occur for sample members who are federal student financial aid applicants.

The initial CPS matching process began after the student sample had been selected for an institution, but before student record (CADE) data collection activities had begun. This matching was against the CPS data for the 2003–04 financial aid year. Since data obtained from CPS were relevant in determining study response status, match rates are presented for all eligible sample members for whom a social security number was available. As shown in table 12, not all sample students were submitted to the CPS for matching. This was primarily because student Social Security numbers and last names were not obtained from some institutions. Following CADE, a number of student cases that had not previously matched successfully to CPS were resubmitted,

based on either a newly obtained Social Security number or the evidence in the institution records that the student had, in fact, applied for federal student aid for the 2003–04 academic year.

The overall matching rate for the 2003–04 CPS data was 60 percent. Match rates varied by type of institution, ranging from 50 percent for public 2-year institutions to 84 percent for private for-profit 2-year institutions.

Approximately 35 percent of graduate/first-professional students matched to the 2003–04 CPS. Also, 64 percent of undergraduate students matched: of these, 69 percent were first-year undergraduates and 59 percent were other undergraduates. Nearly all institutions require undergraduate aid applicants to file a FAFSA in order to determine their eligibility for federal Pell Grants, federal campus-based aid, and federal loans as part of the undergraduate aid packaging process. Graduate/first-professional students are not usually required to file a FAFSA unless they are specifically applying for federal loans, the only type of federal aid generally available to graduate students. Graduate students often apply directly through their institution or department for fellowships and assistantships, which are usually not need-based and do not require the completion of the federal financial aid forms on which CPS matching is based.

The NPSAS:04 sample students were also matched to the 2004–05 CPS files. It was expected that fewer sample students would successfully match to the 2004–05 CPS files, primarily because some students may have completed their postsecondary education during the 2004–05 NPSAS year. Table 12 shows that, overall, 63 percent of sample students matched to either CPS 2003–04 or CPS 2004–05, and 31 percent matched to both data files.

National Student Loan Data System (NSLDS)

Results of the matching to NSLDS loan and Pell Grant files are shown in table 13. Results presented are based only on study respondents since NSLDS data were not required to determine study response status. Successful matching to NSLDS can only occur for sample members who have received federal loans and/or Pell Grants. NSLDS files are historical, thus, information about receipt of such loans and grants was available not only for the NPSAS study year, but also for prior years (where applicable). Therefore, table 13 shows historical match rates for eligible study respondents, which does not necessarily mean that the match was for the current NPSAS year.

In total, 48,840 study respondents (56 percent of those submitted) were matched to the historical loan database. NSLDS match rates ranged from 34 percent for public less-than-2-year institutions, to 87 percent for private for-profit 2-year or more institutions.

Pell Grant matches were obtained for 39,240 study respondents (45 percent of those submitted). The Pell match rate ranged from 27 percent for private not-for profit 4-year doctorate-granting institutions to 79 percent for private for-profit less-than-2-year institutions.

Table 12. Results of Central Processing System (CPS) matching for 2003–04 and 2004–05, by institutional characteristics and student type: 2004

Institutional characteristics and student type[1]	Eligible students[2]	Matched to 2003–04		Matched to 2004–05		Matched to both years		Matched to either year	
		Number[3]	Percent	Number[3]	Percent	Number[3]	Percent	Number[3]	Percent
All students	95,180	57,370	60.3	32,080	33.7	29,080	30.6	60,370	63.4
Institution level									
Less-than-2-year	10,250	8,130	79.3	2,690	26.3	2,570	25.1	8,260	80.5
2-year	34,110	18,770	55.0	10,810	31.7	9,230	27.1	20,350	59.6
4-non-doctorate-granting	19,710	13,900	70.5	8,570	43.5	8,130	41.2	14,340	72.7
4-year doctorate-granting	31,100	16,570	53.3	10,010	32.2	9,150	29.4	17,430	56.0
Institutional control									
Public	60,540	32,020	52.9	18,810	31.1	16,390	27.1	34,440	56.9
Private not-for-profit	21,020	14,000	66.6	8,540	40.6	8,100	38.5	14,440	68.7
Private for-profit	13,620	11,350	83.3	4,740	34.8	4,590	33.7	11,500	84.4
Type of institution									
Public less-than-2-year	2,120	1,370	64.6	520	24.5	460	21.9	1,430	67.2
Public 2-year	29,010	14,460	49.9	8,540	29.5	7,040	24.3	15,970	55.0
Public 4-non-doctorate-granting	8,290	5,240	63.3	3,210	38.7	2,970	35.8	5,480	66.1
Public 4-year doctorate-granting	21,120	10,950	51.8	6,540	31.0	5,920	28.0	11,570	54.8
Private not-for-profit 2-year or less	2,600	2,150	82.5	1,210	46.5	1,160	44.6	2,200	84.4
Private not-for-profit 4-non-doctorate-granting	8,730	6,410	73.4	3,950	45.3	3,790	43.4	6,570	75.3
Private not-for-profit 4-year doctorate-granting	9,690	5,440	56.2	3,380	34.8	3,150	32.5	5,670	58.5
Private for-profit less-than-2-year	7,530	6,260	83.2	1,970	26.1	1,900	25.3	6,320	84.0
Private for profit 2-year or more	6,100	5,090	83.5	2,770	45.4	2,690	44.1	5,170	84.9
Student type									
Total undergraduate	84,190	53,490	63.5	29,990	35.6	27,220	32.3	56,260	66.8
FTB student	37,660	25,860	68.7	14,720	39.1	13,550	36.0	27,040	71.8
Other undergraduate	46,530	27,630	59.4	15,260	32.8	13,670	29.4	29,220	62.8
Graduate/first-professional	10,990	3,870	35.3	2,090	19.1	1,860	16.9	4,110	37.4

[1] Both institutional and student classifications were verified to correct classification errors on the sampling frame.

[2] Includes all eligible students for whom apparently legitimate Social Security numbers were obtained either before or during computer-assisted data entry (CADE).

[3] The number presented reflects the total number of matches of those submitted and may include students who were classified as study nonrespondents.

NOTE: Detail may not sum to totals because of rounding. All percentages are unweighted and based on the number of eligible students within the row under consideration. FTB = first-time beginner.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Table 13. Results of National Student Loan Data System (NSLDS) matching, by institutional characteristics and student type: 2004

Institutional characteristics and student type[2]	Study respondents[3]	Sent to NSLDS		Matched to NSLDS loan[1]		Matched to NSLDS Pell[1]	
		Number	Percent[4]	Number	Percent[4]	Number	Percent[4]
All students	90,750	87,890	96.8	48,840	55.6	39,240	44.6
Institution level							
Less-than-2-year	9,690	9,630	99.4	6,400	66.4	7,330	76.1
2-year	31,260	29,730	95.1	12,610	42.4	14,570	49.0
4-non-doctorate-granting	19,400	18,820	97.0	12,520	66.5	8,700	46.2
4-year doctorate-granting	30,400	29,710	97.7	17,320	58.3	8,640	29.1
Institutional control							
Public	56,990	54,610	95.8	25,070	45.9	21,920	40.1
Private not-for-profit	20,630	20,250	98.1	13,050	64.4	7,840	38.7
Private for-profit	13,120	13,030	99.3	10,720	82.3	9,480	72.8
Type of institution							
Public less-than-2-year	1,930	1,910	99.0	640	33.7	1,190	62.4
Public 2-year	26,320	24,830	94.3	8,740	35.2	11,220	45.2
Public 4-non-doctorate-granting	8,160	7,770	95.3	4,390	56.5	3,450	44.4
Public 4-year doctorate-granting	20,600	20,110	97.6	11,300	56.2	6,060	30.1
Private not-for-profit 2-year or less	2,570	2,500	97.3	1,370	54.6	1,750	69.7
Private not-for-profit 4-non-doctorate-granting	8,550	8,430	98.6	5,880	69.7	3,620	42.9
Private not-for-profit 4-year doctorate-granting	9,510	9,320	98.0	5,810	62.3	2,480	26.6
Private for-profit less-than-2-year	7,150	7,130	99.8	5,600	78.5	5,640	79.1
Private for profit 2-year or more	5,970	5,890	98.7	5,120	87.0	3,840	65.1
Student type							
Total undergraduate	79,850	77,380	96.9	42,620	55.1	37,250	48.1
FTB student	35,510	34,700	97.7	17,270	49.8	16,530	47.6
Other undergraduate	44,340	42,680	96.2	25,350	59.4	20,720	48.5
Graduate/first-professional	10,890	10,510	96.5	6,230	59.2	1,990	19.0

[1] Matching was completed on historical files.
[2] Both institutional and student classifications were verified to correct classification errors on the sampling frame.
[3] Includes all study respondents for whom an apparently legitimate social security number was available.
[4] Percentages are based on the number of eligible students within the row under consideration.

NOTE: Detail may not sum to totals because of rounding. FTB = first-time beginner.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

3.3.2 Outcomes of Student Record Abstraction

As previously indicated, 1,360 of the 1,630 (84 percent) eligible sample institutions provided a student enrollment list that could be used for sample selection (see table 8). These institutions were therefore eligible to participate in the student record abstraction phase of the study referred to as CADE. NPSAS:04 included three abstraction methods for the student record data collection—self-CADE, field-CADE, and data-CADE. Table 14 shows the final data abstraction method for all institutions that completed CADE.

Abstraction method

Of the 1,300 institutions that provided student record data, the majority (66 percent) did so by self-CADE. Data-CADE was the next most common method, with 21 percent of CADE completions being submitted via electronic data files. Field data collectors performed the record abstraction from the remaining 13 percent of CADE completions. Compared to NPSAS:2000, the rate at which institutions opted for the data-CADE in NPSAS:04 was significantly higher: 21 percent compared to 3 percent in NPSAS:2000 ($Z = 12.27$, $p < 0.05$). As was described earlier, student sample sizes were larger than in NPSAS:2000, making the data-CADE option more attractive. Data-CADE was also useful for institutional systems that provided data for students from multiple institutions. There was a corresponding decrease in the use of field-CADE from NPSAS:2000; 13 percent compared to 23 percent ($Z = 6.0$, $p < 0.05$).

Table 14. Student record abstraction method: 2004

CADE abstraction method	Institutions providing CADE		Total students[1]	
	Number	Percent[2]	Number	Percent[2]
Total	1,300	100.0	103,620	100.0
Abstraction method				
Self-CADE	860	65.8	48,860	47.2
Data-CADE	280	21.1	33,210	32.0
Field-CADE	170	13.1	21,550	20.8

[1] The total represents the number of students sampled from institutions that completed computer-assisted data entry (CADE) and may include students who were classified as study nonrespondents.
[2] Percentage of total number of eligible institutions/students.
NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

CADE completion rates

At the institution level, an institution was classified as having completed CADE if sufficient data were obtained for at least one sample student. Institution-level weighted and unweighted CADE completion rates are shown in Table 15. Overall, 96 percent (weighted) of the participating institutions (those that provided enrollment lists from which a student sample could be selected) completed CADE.

A student record was considered to represent a CADE record "complete" if it had nonmissing data for any one or more of the following critical items:

- received financial aid;
- enrollment;
- tuition;
- degree program; and
- race.

Table 15. Institutional-level computer-assisted data entry (CADE) completion rates, by institutional characteristics and abstraction method: 2004

Institutional characteristics and abstraction method[1]	Institutions providing lists	Institutions providing CADE		
		Number	Unweighted Percent[2]	Weighted Percent
Total	1,360	1,300	95.4	96.3
Institutional level				
Less-than-2-year	200	190	94.0	94.0
2-year	410	390	95.3	96.9
4-year non-doctorate-granting	380	360	95.8	96.5
4-year doctorate-granting	380	360	95.8	95.6
Institutional control				
Public	680	660	96.9	96.2
Private not-for-profit	450	430	94.3	96.8
Private for-profit	230	210	93.0	94.5
Type of institution				
Public less-than-2-year	50	50	100.0	100.0
Public 2-year	310	310	98.1	97.2
Public 4-year nondoctorate-granting	110	110	96.5	96.4
Public 4-year doctorate-granting	200	190	94.5	94.8
Private not-for-profit 2-year or less	60	50	81.0	82.7
Private not-for-profit 4-year non-doctorate granting	220	210	95.8	97.0
Private not-for-profit 4-year doctorate granting	180	170	97.2	97.4
Private for-profit less than-2-year	130	120	93.0	93.1
Private for profit 2-year or more	100	90	93.0	95.0
Abstraction method				
None chosen	10	†	†	†
Self-CADE	910	860	94.5	93.8
Data-CADE	280	280	99.3	99.0
Field-CADE	170	170	97.7	97.5

† Not applicable.
[1] Institutional characteristics were verified (where possible) to correct classification errors on the sample frame.
[2] Percentage of institutions providing lists.
NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Completion rates ranged from 94 percent (weighted) for institutions choosing self-CADE to 99 percent for data-CADE. CADE completion rates varied by type of institution, ranging from 83 percent from private not-for-profit 2-year or less institutions to 100 percent for public less-than-2-year institution.

Student-level CADE completion rates are presented in table 16 by type of institution and student type. Overall, the student-level CADE completion rate (the percentage of study-eligible cases for whom a completed CADE record was obtained) was 92 percent (weighted). Weighted student-level completion rates ranged from 71 percent for private not-for-profit 2-year or less institutions, to 96 percent for public less-than-2-year institutions. Weighted completion rates by student type were about 92 percent for undergraduate and 93 percent for graduate and first-professional students.

Table 16. Student-level computer-assisted data entry (CADE) completion rates, by institutional characteristics and student type: 2004

Institutional characteristics and student type[2]	Eligible students[3]	CADE completes[1]		
		Number	Unweighted percent	Weighted percent
Total	101,010	88,920	88.0	91.7
Institutional level				
Less-than-2-year	10,330	8,800	85.2	87.8
2-year	37,750	32,150	85.2	88.9
4-year non-doctorate-granting	20,630	18,530	89.8	94.4
4-year doctorate-granting	32,310	29,440	91.1	93.7
Institutional control				
Public	65,540	58,400	89.1	91.6
Private not-for-profit	21,660	18,920	87.3	93.2
Private for-profit	13,820	11,600	84.0	89.2
Type of institution				
Public less-than-2-year	2,150	2,020	93.9	95.9
Public 2-year	32,540	28,580	87.8	89.6
Public 4-year non-doctorate-granting	8,890	7,900	88.9	94.8
Public 4-year doctorate-granting	21,960	19,910	90.6	93.2
Private not-for-profit 2-year or less	2,730	1,700	62.2	70.8
Private not-for-profit 4-year non-doctorate granting	8,880	7,990	89.9	93.1
Private not-for-profit 4-year doctorate granting	10,050	9,240	91.9	94.7
Private for-profit less than-2-year	7,550	6,350	84.1	87.0
Private for profit 2-year or more	6,260	5,250	83.8	90.2
Student type				
Total undergraduate	89,460	78,590	87.9	91.5
Potential FTB	39,440	34,590	87.7	90.5
Other undergraduates	50,020	44,000	88.0	92.0
Graduate/first professional	11,560	10,340	89.4	93.0

[1] Eligible students who met the criteria for qualification as a CADE completion, which required an indication of financial aid receipt, enrollment status, tuition, degree program, or race in the CADE instrument. Numbers presented here may include students who were classified as study nonrespondents.

[2] Both institutional characteristics and student classifications were verified (where possible) to correct classification errors on the sample frame.

[3] Students determined to be eligible in CADE and/or the student interview.

NOTE: Detail may not sum to totals because of rounding. FTB = first-time beginner.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

3.3.3 Student Interview Completion

Locating

When dealing with a mobile group such as the NPSAS:04 student sample, locating can be one of the more difficult tasks. A variety of approaches were used during NPSAS:04 to locate and interview sampled students. These approaches included the use of an initial mailing to all students, follow-up letters and e-mails to nonrespondents, telephone tracing (calling local and permanent numbers as well as any other numbers obtained during the course of contacting), and intensive tracing (i.e., using consumer databases, Web searches, and a variety of directories).

As shown in table 17, of the 101,010 eligible sample members, 79 percent were successfully located. The highest location rates were for students attending public 4-year doctorate-granting institutions (86 percent), while the lowest location rates were among those from private for-profit less-than-2-year institutions (66 percent) ($\chi^2 = 2{,}506$, $p < 0.001$). Graduate students proved the easiest group to find, with 88 percent of these students being located, compared to 77 percent of other undergraduates, and 80 percent of FTB undergraduates ($\chi^2 = 684$, $p < 0.001$).

Table 17. Student locating, by institutional characteristics and student type: 2004

Institutional characteristics and student type[1]	Total	Located	
		Number	Percent
Total	101,010	80,050	79.2
Institutional level			
Less-than-2-year	10,330	7,030	68.0
2-year	37,750	28,210	74.7
4-non-doctorate-granting	20,630	17,130	83.0
4-year doctorate-granting	32,310	27,690	85.7
Institutional control			
Public	65,540	52,360	79.9
Private, not-for-profit	21,660	18,140	83.7
Private, for-profit	13,820	9,550	69.2
Type of institution			
Public less-than-2-year	2,150	1,650	76.7
Public 2-year	32,540	24,540	75.4
Public 4-non-doctorate-granting	8,890	7,370	82.9
Public 4-year doctorate-granting	21,960	18,800	85.6
Private not-for-profit 2-year-or-less	2,730	1,930	70.5
Private not-for-profit 4-non-doctorate-granting	8,880	7,580	85.4
Private not-for-profit 4-year doctorate-granting	10,050	8,640	85.9
Private for-profit less-than-2-year	7,550	4,950	65.6
Private for-profit 2-year-or-more	6,260	4,600	73.5
Student type			
Total undergraduate	89,460	69,900	78.1
FTB student	39,440	31,430	79.7
Other undergraduate	50,020	38,470	76.9
Graduate/first-professional	11,560	10,150	87.8

[1] Both institutional and student classifications were verified to correct classification errors on the sampling frame.

NOTE: Detail may not sum to totals because of rounding. Excludes 8,200 cases determined to be ineligible for the study. FTB = first-time beginner.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Table 18 presents the results of matching to the various batch searches used to obtain locating information for sample members (described in chapter 2). Telematch was the most successful, with 50 percent of cases returning address information. The National Change of Address (NCOA) system and FastData returned locating information on 9 percent and 6 percent, respectively, of the cases submitted.

Table 18. Batch processing record match rates, by tracing source: 2004

Method of tracing	Number of records sent	Number of records matched	Percent matched[1]
Total	240,750	65,060	27.0
NCOA	109,210	9,360	8.6
Telematch	109,210	54,390	49.8
FastData	22,330	1,310	5.9

[1] Percent is based on the number of records sent for batch tracing. Since records were sent to multiple tracing sources, multiple record matches were possible.

NOTE: NCOA = National Change of Address.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04)

Intensive tracing during data collection

Intensive tracing efforts were required for cases in which no interview was obtained via self-administration nor did the preloaded computer-assisted telephone interview (CATI) locating information result in contact with the sample member. These cases were assigned to RTI Call Center Services' (CCS) Tracing Services for intensive centralized tracing, utilizing searches of public and proprietary databases, the Web, and a variety of information directories. Overall, one-fourth (26 percent) of eligible sample members required intensive tracing efforts (table 19). Intensive tracing varied by institution type, ranging from 17 percent for private not-for-profit 4-year doctorate-granting institutions, to 38 percent for private for-profit less than 2-year institutions. Intensive tracing also varied by student type: 19 percent for graduate and first-professional students, and 27 percent for undergraduate students.

Table 19. Students requiring intensive tracing procedures, by institutional characteristics and student type: 2004

Institutional characteristics and student type[1]	Total	Cases requiring intensive tracing efforts	
		Number	Percent
Total	101,100	25,940	25.7
Institutional level			
Less-than-2-year	10,350	3,730	36.0
2-year	37,780	11,920	31.5
4-non-doctorate-granting	20,640	4,350	21.1
4-year doctorate-granting	32,320	5,940	18.4
Institutional control			
Public	65,590	16,950	25.8
Private, not-for-profit	21,670	4,240	19.6
Private, for-profit	13,840	4,750	34.3
Type of institution			
Public less-than-2-year	2,150	630	29.2
Public 2-year	32,570	10,260	31.5
Public 4-year non-doctorate-granting	8,890	1,910	21.4
Public 4-year doctorate-granting	21,970	4,160	18.9
Private not-for-profit 2-year-or-less	7,570	2,870	37.9
Private not-for-profit 4-year non-doctorate-granting	8,880	1,680	18.9
Private not-for-profit 4-year doctorate-granting	10,060	1,740	17.3
Private for-profit less-than-2-year	7,570	2,870	37.9
Private for-profit 2-year-or-more	6,270	1,880	30.0
Student type			
Total undergraduate	89,540	23,780	26.6
FTB student	39,490	10,170	25.8
Other undergraduate	50,050	13,610	27.2
Graduate/first-professional	11,560	2,160	18.7

[1] Both institutional and student classifications were verified to correct classification errors on the sampling frame.

NOTE: Detail may not sum to totals because of rounding. Excludes 8,200 cases determined to be ineligible for the study. FTB = first-time beginning.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Table 20 show that of the 25,940 eligible cases requiring intensive tracing, 10,870 (42 percent) were ultimately located, and approximately 30 percent of them were interviewed.

Table 20. Locate and interview rates, by intensive tracing efforts: 2004

		Located		Interviewed		
	Total	Number	Percent	Number	Percent	Weighted percent
Total	101,100	80,090	79.2	62,220	61.5	69.7
Intensive tracing required	25,940	10,880	41.9	7,850	30.3	77.2
No intensive tracing required	75,160	69,220	92.1	54,370	72.3	42.5

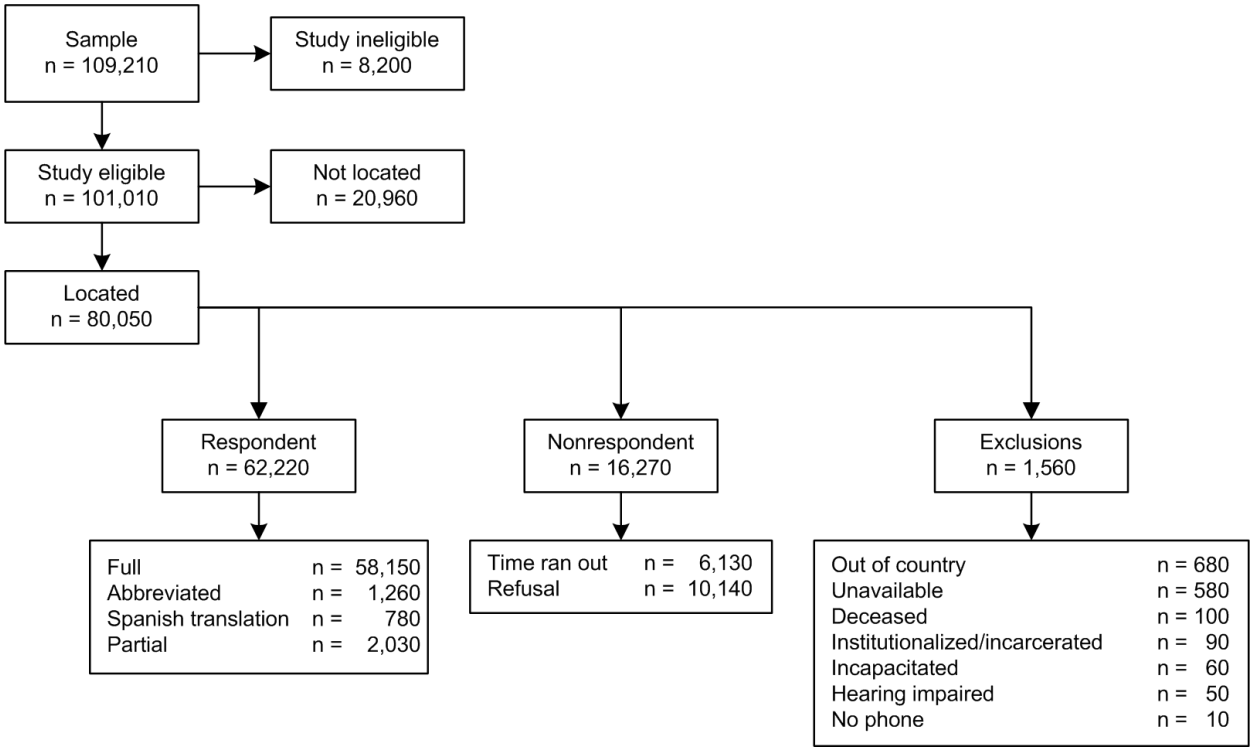
NOTE: Detail may not sum to totals because of rounding. Excludes 8,200 cases determined to be ineligible for the study.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

3.3.4 Student Locating and Response Rate Summary

Overall locating and interviewing outcomes are shown in figure 9. Of the 109,210 sample members, 80,050 (73 percent) were located, 20,960 (19 percent) were not located, and 8,200 (8 percent) were located but determined to be ineligible for the study. Of the located sample members, 78 percent completed either a full interview, an abbreviated interview used to capture critical information from students with a high probability of nonresponse, a hardcopy Spanish interview or completed enough of the questionnaire to be considered a partial interview.[32]

Figure 9. NPSAS:04 locating and interview outcomes



NOTE: Detail may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Table 21 presents student interview completion rates among eligible sample members by institutional characteristics and student type. The weighted response rate for the student data interview was 71 percent. Weighted student interview completion rates ranged from 49 percent for private-for-profit less-than-2-year institutions, to 74 percent for 4-year doctorate-granting institutions (public and private, not-for-profit). Weighted completion rates by student type were 72 percent for undergraduates and 75 percent for graduate and first-professional students.

[32] Students who completed the enrollment section of the questionnaire but did not complete the entire survey were considered partial interviews.

Table 21. Student interview completion results, by institutional characteristics and student type: 2004

Institutional characteristics and student type[2]	Eligible students[3]	Completed interviews[1]		
		Number	Unweighted percent	Weighted percent
Total	99,450	62,220	62.6	70.6
Institutional level				
Less-than-2-year	10,210	4,830	47.3	50.2
2-year	37,130	20,790	56.0	69.3
4-non doctorate-granting	20,340	13,840	68.0	70.8
4-year doctorate-granting	31,770	22,760	71.6	73.9
Institutional control				
Public	64,520	40,620	63.0	71.3
Private, not-for-profit	21,290	14,620	68.7	71.8
Private, for-profit	13,640	6,970	51.1	60.4
Type of institution				
Public less-than-2 year	2,130	1,200	56.4	61.6
Public 2-year	31,990	18,000	56.3	69.8
Public 4-non-doctorate-granting	8,760	5,890	67.2	71.9
Public 4-year doctorate-granting	21,640	15,530	71.8	73.8
Private not-for-profit 2-year or less	2,690	1,350	50.3	56.3
Private not-for-profit 4-non-doctorate-granting	8,760	6,250	71.3	70.7
Private not-for profit 4-year doctorate-granting	9,840	7,030	71.4	74.0
Private for-profit less than 2-year	7,470	3,420	45.8	48.6
Private for-profit 2-year or more	6,170	3,550	57.5	65.7
Student type				
Total undergraduate	88,030	53,680	61.0	71.9
FTB student	38,850	25,030	64.4	77.4
Other undergraduate	49,180	28,650	58.3	66.3
Graduate/first professional	11,420	8,540	74.8	75.1

[1] Eligible students who met the criteria for qualification as a student interview completion, which required completing at least a partial interview.
[2] Both institutional and student classifications were verified to correct classification errors on the sampling frame.
[3] Excludes 8,200 cases determined to be ineligible for the study and 1,560 cases who were either deceased, unavailable for the duration of the survey, out of the country, incapable/incapacitated, institutionalized/incarcerated, had no phone, or were hearing impaired.

NOTE: Detail may not sum to totals because of rounding. FTB = first-time beginner.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

As was described in chapter 2, data collection notifications were sent to all sample members, inviting them to participate by completing the web-based self-administered interview. Sample members were given 4 weeks to complete the interview, during which time e-mail reminders were sent to cases for whom we had an e-mail address. After the 4-week period, outbound telephone interviewing began. However, sample members were always encouraged to complete the self-administered interview at their convenience.

Completion mode for student interviews is presented in table 22. Among the 62,220 completed student interviews, 28 percent (weighted) were completed via self-administration during the first 4 weeks after notification. Fifty-three percent of completed student interviews were conducted with telephone interviewers, and the remaining 19 percent were completed via self-administration after the early incentive period had expired.

Table 22. Student interview completion mode: 2004

	Completed interviews		
	Number	Unweighted percent	Weighted percent
Total	62,220	100.0	100.0
Self-administered	28,710	46.1	46.7
Self-administered: early response period	17,100	27.5	27.5
Self-administered: with prompting	11,610	18.7	19.2
Interviewer-administered	33,510	53.9	53.3

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

3.3.5 Conversion of Nonrespondents

As described earlier, all sample members were invited to participate in the student interview. Those who did so within the first 4 weeks were offered an incentive. Following the initial 4-week period, data collection continued with telephone prompting, and no offer of incentive. Once cases were identified as nonrespondents, additional mailings and e-mail prompts were used in conjunction with incentives to encourage participation in NPSAS:04. Letters for each mailing contained the same general information but were tailored to the type of nonrespondent (e.g., refusal, hard to reach cases, etc.). (See appendix C for materials sent to the sample members.) Letters, e-mails, and subsequent telephone prompts offered respondents a monetary incentive for completing the interview.

Refusal conversion letters were sent on a flow basis to sample members who initially refused to participate in the study. These letters were tailored to address the typical concerns expressed by those refusing to participate. In all, 11,840 students were sent a refusal letter and 9,320 students were sent an e-mail message containing the same information as the letter. Of the 22,620 eligible students identified as refusals (either by the sample member or someone else), 8,270 were interviewed (37 percent).

Another letter was tailored for use with nonrespondents who did not actively refuse to participate, e.g., those for whom 20 or more call attempts had been made, but an interview had not been completed. In all, 52,930 students were sent a nonresponse letter and 38,060 students were sent an e-mail. Of the 50,070 eligible students identified as nonrespondents, 19,480 were interviewed (39 percent).

Approximately 2 weeks before the end of the data collection period, all nonrespondents (refusals and nonrefusals alike) were sent a final mailing and/or e-mail asking for their participation. Of the 40,950 eligible students that were sent the end-of-study letter or e-mail, 9,070 (22 percent) were ultimately interviewed. A smaller group of respondents (6,890) were sent a final request for participation via a postcard. Of the 6,670 eligible students that were sent the end-of-study postcard, 2,720 (41 percent) were ultimately interviewed.

3.4 Completeness of Data Records among Study Respondents

As discussed in section 3.2, a study respondent is defined as any eligible student for whom sufficient data were obtained from one or more sources. The sources used to define study response status include institutional records, student interview, and the Department of Education's CPS. The completeness of data records across sources among study respondents is presented in table 23. In addition to the three sources used to determine the study respondents, NSLDS loan and Federal Pell Grant data are also included in the table. Like CPS, these sources are used to supplement the institutional record and student interview data.

In total, 92 percent (weighted) of the study respondents have student record data from the NPSAS institution (CADE data). The percentage of study respondents who have student interview data is 70 percent. Additionally, 52 percent of study respondents had a federal aid application for the 2003–04 academic year in the CPS database. The percentage of study respondents who matched to the NSLDS loan database for the 2003–04 academic year is 34 percent. Those that matched to the NSLDS Federal Pell Grant database for the same year is 23 percent.

Table 23. Percent of student respondents with data, by institutional characteristics, student type, and source: 2004

Institutional characteristics and student type[5]	Number of responding students[6]	Student record percent[1]		Interview percent[2]		CPS percent[3]		NSLDS loans Percent[7]		NSLDS Pell Grants percen[4]	
		Un-weighted	Weighted	Un-weighted	Weighted	Un-weighted	Weighted	Un-weighted	Weighted	Un-weighted	Weighted
Total	90,750	90.2	91.7	68.5	69.7	62.7	52.4	37.7	33.8	32.0	23.1
Institutional level											
Less-than-2-year	9,690	87.6	87.8	49.8	49.6	83.3	69.9	44.5	47.6	63.8	47.7
2-year	31,260	87.4	88.9	66.3	68.5	59.1	43.1	24.3	16.5	36.1	24.7
4-year non-doctorate-granting	19,400	91.7	94.4	71.3	70.0	71.2	63.0	51.5	48.7	33.5	26.7
4-year doctorate-granting	30,400	92.8	93.7	74.8	72.8	54.4	55.0	40.4	43.5	16.8	16.4
Institutional control											
Public	56,990	91.3	91.6	71.1	70.5	55.5	46.8	28.2	25.6	27.3	21.5
Private not-for-profit	20,630	89.2	93.2	70.8	70.8	67.7	62.5	49.0	50.7	27.6	19.5
Private for-profit	13,120	86.9	89.2	53.1	59.8	86.0	82.5	61.2	71.8	59.5	47.4
Type of institution											
Public less than-2-year	1,930	94.6	95.9	62.1	60.6	71.1	34.1	14.9	11.0	50.4	21.0
Public 2-year	26,320	90.5	89.6	68.2	69.0	53.9	40.5	17.3	13.0	32.1	22.5
Public 4-year nondoctorate-granting	8,160	90.8	94.8	72.2	71.1	63.2	55.8	40.0	38.4	30.8	25.0
Public 4-year doctorate-granting	20,600	92.2	93.2	75.4	72.9	53.0	53.2	38.6	40.6	17.6	18.2
Private not-for-profit 2-year or less	2,570	64.9	70.8	52.5	55.6	83.1	77.4	41.0	45.1	58.8	46.6
Private not-for-profit 4-year non-doctorate granting	8,550	91.2	93.1	73.0	69.8	74.8	65.6	56.7	52.6	32.2	25.0
Private not-for-profit 4-year doctorate granting	9,510	93.9	94.7	73.8	72.7	57.1	58.1	44.2	48.9	15.1	11.7
Private for-profit less than-2-year	7,150	87.3	87.0	47.9	48.1	86.8	76.1	54.9	55.0	66.8	51.7
Private for profit 2-year or more	5,970	86.3	90.2	59.3	65.0	85.2	85.4	68.6	79.2	50.8	45.4
Student type											
Total undergraduate	79,850	90.0	91.5	67.1	69.0	66.4	53.8	38.7	33.0	36.3	
Potential FTB	35,510	89.7	90.5	70.4	76.6	72.1	59.0	38.9	31.3	40.6	30.9
Other undergraduates	44,340	90.2	92.0	64.5	65.4	61.9	51.3	38.5	33.8	32.9	24.3
Graduate/first professional	10,890	91.6	93.0	78.3	74.5	35.5	43.0	30.2	39.0	0.4[7]	0.7[7]

[1] Percent of study respondents who met the criteria for qualification as a computer-assisted data entry (CADE) completion.

[2] Percent of study respondents who met the criteria for qualification as a student interview completion.

[3] Percent of study respondents who matched to CPS, which contains federal aid application (FAFSA) data.

[4] Percent of study respondents who matched to the National Student Loan Data System (NSLDS) for loans and Pell Grants during the 2003–04 academic year.

[5] Both institutional characteristics and student classifications were verified (where possible) to correct classification errors on the sample frame.

[6] A responding student is defined as any eligible student for whom sufficient data were obtained from one or more sources, including: student interview, institutional records, and the Department of Education's Central Processing System (CPS).

[7] The small percentage of matched graduate and first-professional study respondents were undergraduates at some time during the year and as such were eligible for this type of aid during the year.

NOTE: Detail may not sum to totals because of rounding. FTB = first-time beginner.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

This page left intentionally blank.

Chapter 4
Evaluation of Field Operations and Data Quality

Evaluation of study methodology and procedures, as well as of study outcomes, were planned and conducted throughout the course of the 2004 National Postsecondary Student Aid Study (NPSAS:04). The results of these quantitative and qualitative analyses provide information pertaining to the efficacy of study data and are also useful in planning for subsequent waves of NPSAS.

4.1 Enrollment List Collection

4.1.1 Early Contacting Activities

Making early contact with institutions was an important part of the design of NPSAS:04. The scheduled release of data required an accelerated data collection schedule, which required that enrollment lists were received in time to allow for sampling, student interviewing, and data processing to be completed by December 2004. As such, much focus was devoted to the activities of institutional early contacting.

Table 24 presents the flow of enrollment list receipt in NPSAS:96[33] and NPSAS:04. The 1,360 lists received by July 2004 provided a sufficiently large and representative student sample to allow list collection to end. The flow of list receipt was very similar for both studies.

Table 24. Cumulative flow of enrollment list receipt: 1996 and 2004

	Cumulative percentage of lists received	
Month	NPSAS:96	NPSAS:04
1	17.7	12.5
2	42.2	38.4
3	63.6	58.8
4	85.1	75.4
5	95.9	88.7
6	98.8	98.2
7	100.0	100.0

SOURCE: U.S. Department of Education, National Center for Education Statistics,1996 National Postsecondary Student Aid Study (NPSAS:96), and 2004 National Postsecondary Student Aid Study (NPSAS:04).

4.1.2 Institutional Participation

Institutional participation was evaluated for potential effects of prior NPSAS participation. Summary results of these analyses are shown in table 25. Among eligible institutions, the NPSAS:04 enrollment list provision rate among the 980 institutions that had previously participated in NPSAS was 84 percent, which is not statistically different than the rate among institutions that had not previously participated (83 percent; $\chi^2 = 0.18$, $p > 0.05$).

[33] NPSAS:96 is used for this comparison because it was the most recent study from which the BPS cohort identified.

Table 25. Institutional NPSAS:04 enrollment list participation, by prior NPSAS participation

Institutional characteristics[1]	Eligible institutions	No prior NPSAS participation			Participated at least once		
			Provided lists			Provided lists	
		Number	Number	Percent[2]	Number	Number	Percent[3]
All institutions	1,630	650	540	83.1	980	830	83.8
Institution level							
Less-than-2-year	240	170	150	85.9	70	50	77.9
2-year	480	240	200	83.9	240	200	86.0
4-year non-doctorate-granting	460	180	150	81.0	270	230	84.6
4-year doctorate-granting	460	50	40	77.4	410	340	83.1
Institutional control							
Public	800	240	200	85.0	560	480	84.9
Private not-for-profit	560	210	170	80.3	350	280	81.8
Private for-profit	270	200	160	83.7	80	70	85.5
Type of institution							
Public less than-2-year	70	40	30	85.0	30	20	74.1
Public 2-year	370	170	140	85.7	200	170	84.1
Public 4-year non-doctorate-granting	140	20	20	79.2	110	100	85.6
Public 4-year doctorate-granting	230	10	10	87.5	220	190	86.5
Private not-for-profit 2-year or less	70	40	40	85.7	30	30	96.4
Private not-for-profit 4-year non-doctorate-granting	260	130	100	79.7	140	110	83.1
Private not-for-profit 4-year doctorate-granting	230	40	30	76.7	180	140	78.6
Private for-profit less than-2-year	150	120	100	86.4	30	30	79.4
Private for profit 2-year or more	120	80	60	79.5	40	40	90.5

[1] Institutional classifications were verified by the institutions to correct classification errors on the sampling frame.
[2] Percents are based on the count of eligible institutions with no prior NPSAS participation within the row under consideration.
[3] Percents are based on the count of eligible institutions with prior NPSAS participation within the row under consideration.

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Institutional participation was also examined in terms of the 2000 Carnegie classification categories, as shown in table 26. Table 27 shows the number of historically Black colleges and universities (HBCUs) participating in the current and prior NPSAS rounds.

Table 26. Distribution of participating NPSAS:04 institutions, by 2000 Carnegie classification

Carnegie institutional classification (2000)	Number	Percent
All institutions	1,360	100.0
Doctorate-granting/research extensive	130	9.5
Doctorate-granting/research intensive	90	6.9
Master's I	180	13.0
Master's II	20	1.7
Bachelor's I	60	4.1
Bachelor's II	80	6.0
Bachelor/associate's colleges	10	0.8
Associate's colleges	360	26.4
Theological	30	2.2
Medical	40	2.6
Other health	10	0.7
Engineering and technology	20	1.2
Business and management	10	0.7
Other[1]	40	2.6
Not classified	300	21.6

[1] Includes art/music/design, law, teaching, other specialized, and tribal colleges and universities.

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Table 27. NPSAS participation of historically Black colleges and universities (HBCUs): 1987–2004

NPSAS participation	Number of HBCUs participating	HBCUs as a percent of total number of participating institutions
NPSAS:87	20	1.9
NPSAS:90	20	1.5
NPSAS:93	30	2.6
NPSAS:96	20	1.9
NPSAS:2000	20	2.3
NPSAS:04	30	2.1

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

4.1.3 Quality of Enrollment Lists

Although an electronic list was preferred, institutions were informed that they could provide lists in their preferred format. Of all participating institutions, about 98 percent of institutions provided some type of electronic list, and the remaining 2 percent sent paper lists.

Once lists were received, they were evaluated in terms of appropriateness of format and documentation (relative to instructions provided), as well as for the accuracy of student counts (see chapter 2 for a description of quality control procedures). Table 28 presents the major types of discrepancies encountered. About 44 percent of the institutions provided lists with one or more such problems. The most common problem was that enrollment counts were out of bounds when compared with the Integrated Postsecondary Education Data System (IPEDS) (about 35 percent). The check was not suspended or relaxed (unlike some prior rounds of NPSAS) because

many of the institutions that were called about the discrepancy indicated that the sampling list counts were, in fact, incorrect.

In the event that an enrollment list failed the quality control check, RTI staff contacted the institution to resolve the problem or obtain a new list. After any necessary revisions, all but two lists[34] submitted were usable for selecting the student sample.

Table 28. Types of discrepancies encountered with student lists, by highest level of offering: 2004

Type of institution	Number of institutions	Type of discrepancy encountered[1]	Number	Percent[2]
All institutions	1,360	None	760	55.9
		Count out of bounds	470	34.8
		Unreadable file/list	#	0.3
		Could not identify strata	40	2.6
		Insufficient documentation	20	1.6
		Multiple problems	70	4.8
Less-than-2-year	200	None	110	56.8
		Count out of bounds	60	31.2
		Unreadable file/list	#	0.5
		Could not identify strata	10	5.0
		Insufficient documentation	#	1.5
		Multiple problems	10	5.0
2-year	400	None	210	51.6
		Count out of bounds	170	42.5
		Unreadable file/list	#	0.2
		Could not identify strata	10	1.2
		Insufficient documentation	#	1.0
		Multiple problems	10	3.5
4-year non-doctorate-granting	380	None	240	61.8
		Count out of bounds	110	29.2
		Unreadable file/list	#	0.3
		Could not identify strata	10	2.6
		Insufficient documentation	10	1.3
		Multiple problems	20	4.7
4-year doctorate-granting	380	None	210	54.2
		Count out of bounds	130	33.9
		Unreadable file/list	#	0.3
		Could not identify strata	10	2.6
		Insufficient documentation	10	2.6
		Multiple problems	20	6.3

Rounds to zero.

[1] Categories are mutually exclusive, with an institution being included in only one category within highest level of offering.

[2] Percents are based on the number of institutions within each institution type.

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

[34] These institutions were classified as nonparticipants.

4.2 Student Record Abstraction

Procedures to abstract information from institutional student records (computer-assisted data entry [CADE]) were first implemented in NPSAS:93. Over the years, the procedures have improved for each round of the study to enhance the effectiveness and user-friendliness of the approach, particularly for institutional staff. Most notably, these include the web-based CADE system (web-CADE) used for self-administration by institutional staff and by field interviewers, and the option of submitting data via electronic files (data-CADE).

Other CADE procedures were used to facilitate the timeliness of CADE completion. These included (1) maintaining a help desk to resolve operational or interpretational problems, (2) scheduling calls to prompt self-CADE and data-CADE institutions to complete data abstraction and to answer questions that may have arisen, (3) prescheduling institutions for field staff, and (4) scheduling weekly conferences with field staff to assess their progress.

4.2.1 Preloading Data into CADE

To reduce the data entry effort associated with institutional student record abstraction, certain elements were preloaded into CADE records prior to collection at the institution. Table 29 summarizes the nature and source of preloaded data elements. This included customizing the financial aid award section of CADE to include nonfederal aid that was common to a particular institution. Such customization proved highly successful during NPSAS:96 and NPSAS:2000, and was continued for NPSAS:04.

Table 29. Nature and source of elements preloaded into computer-assisted data entry (CADE): 2004

CADE data element set	Data source
Institution name/ID	IPEDS
Names of most common state financial aid awards	NASSGAP report
Names of most common institution financial aid awards	Institutional coordinator
Institution clock/credit hour indicator	Institutional coordinator
Institution term names and dates	Institutional coordinator
GPA scale	Institutional coordinator
Student name, SSN from institutional records	Enrollment list
Student type indicator (undergraduate/graduate/first-professional)	Enrollment list
Student local and permanent addresses	Enrollment list
Student date of birth, veteran status, and citizenship	CPS record
Student address, phone number, driver's license number and state	CPS record
Student dependency and expected family contribution	CPS record
Flag indicating whether or not student matched to CPS	CPS record

NOTE: IPEDS = Integrated Postsecondary Education Data System; GPA = grade-point average; SSN = Social Security number; CPS = Central Processing System.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Data were preloaded from a variety of sources. These sources include IPEDS and the National Association of State Student Grant and Aid Programs (NASSGAP) state aid report, in addition to data collected from contact with the institutional coordinator and from enrollment lists. The most extensive set of preloaded data were obtained from the Central Processing System (CPS) for federal financial aid applicants. The data from the CPS were used in two different ways. Some items were prefilled with the data from the CPS and users could simply leave it

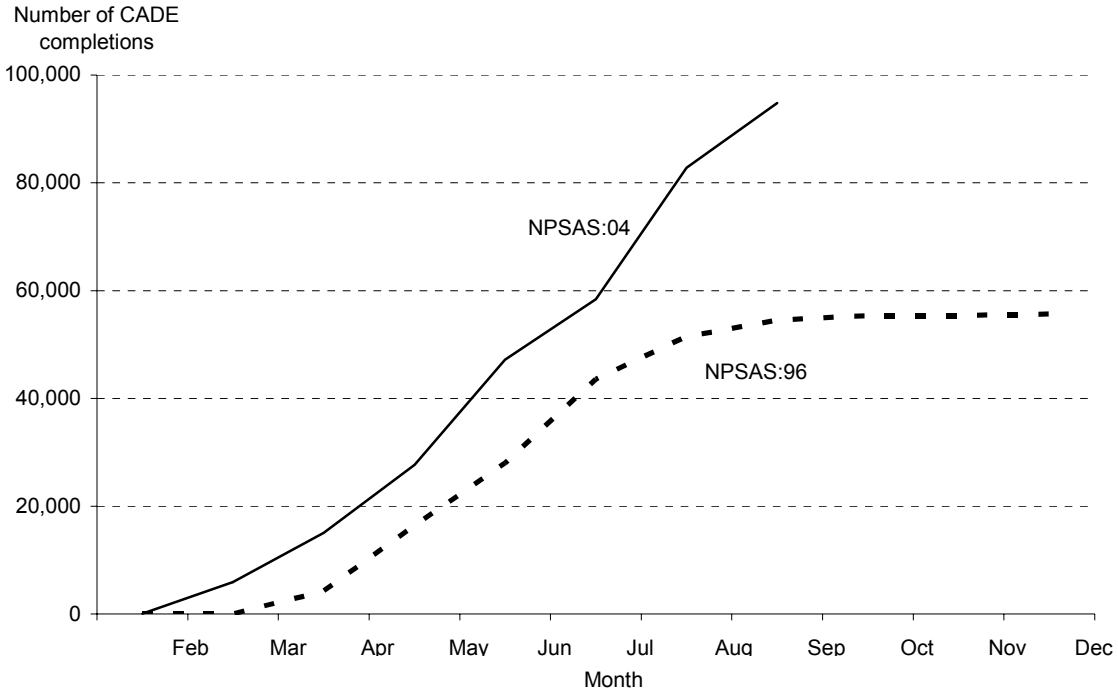
there if it was correct. These data elements included the student's address, phone number, driver's license number, driver's license state, dependency status, and expected family contribution to postsecondary education costs. Other items were preloaded to validate the data entered by users. If users entered something different from what was preloaded from CPS, they would get a warning indicating the difference and could choose to accept the data from CPS or to keep the data originally entered. These variables included citizenship status, veteran status, and student date of birth.

4.2.2 Timeliness of Record Abstraction

CADE systems were prepared on an institution-by-institution basis as enrollment lists were received, samples selected, and matching to CPS was completed. Institutions that opted to provide data via self-CADE began receiving notification that their systems had been initialized in mid-February 2004. An e-mail was also sent to the institutional coordinator informing them that a packet had been mailed and providing them with their username and password to begin accessing the secured website. The first set of field-CADE data collectors began record abstraction activities in April 2004. Final data-CADE specifications and systems for uploading files were also available to institutions in April, with the first successful loading of data files occurring in May. Initialization of CADE systems continued through July 2004.

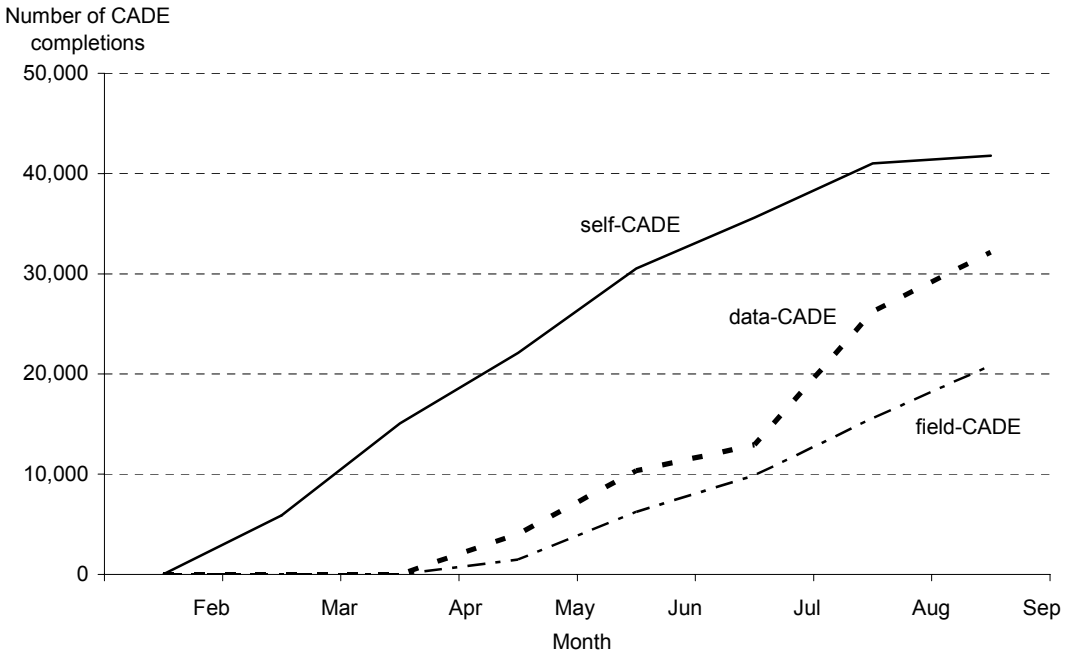
Figure 10 shows the flow of CADE completions, comparing NPSAS:96 and NPSAS:04. Although NPSAS:04 CADE data collection was more condensed than NPSAS:96 CADE data collection, data were collected on many more cases in a shorter time period. The success of early institutional contacting enabled an earlier initialization of CADE data collection. Figure 10 also shows that NPSAS:04 experienced an increase in the number of CADE completions cases in late summer. This increase was primarily due to the large number institutions completing via data-CADE, which can be seen in figure 11. Data-CADE was used largely by institutional systems that provided data files for multiple institutions. Both self-CADE and field-CADE experienced a relatively steady flow of completed cases.

Figure 10. Cumulative flow of computer-assisted data entry (CADE) completions: 1996 and 2004



SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Figure 11. Computer-assisted data entry (CADE) completions, by abstraction mode: 2004



SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

4.2.3 CADE Data Completeness

As discussed in section 3.3.2, a student-level CADE completion required nonmissing data for any one or more of the following critical items:

- receipt of financial aid;
- enrollment;
- tuition;
- degree program; or
- race.

Under this definition, 92 percent (weighted) of the eligible sample students were classified as CADE completes (see table 17). Of the 88,920 CADE completes, 81,810 (92 percent) were determined to be study respondents. The following evaluation presents results for study respondents only.

Table 30 presents item-level completion rates for key data elements among CADE completes overall and by mode of abstraction. It is not surprising that item-level response rates differ among data elements, since institutional record-keeping systems vary dramatically. Not all data elements are available at every institution. However, most of the key data elements showed a high percentage of item-level completeness.

Overall, item-level response rates were very high. Two items had high rates of missing data: marital status and additional phone numbers. Student records frequently lack these items. Response rates varied somewhat by mode of abstraction; in general, data-CADE showed the highest rates of missing data. With the exception of veteran status and phone numbers, self-CADE had higher item-level completion rates than those completed in field-CADE. Data-CADE experienced the lowest item-level completion rates for all but three items (Hispanic status, student class level, and financial aid.) Both self-CADE and field-CADE utilized online edit checks and verifications. This feature is not available for data-CADE, which may have contributed to the higher rate of missing data for this mode.

Table 30. Comparison of NPSAS:04 student record data element completion rates, by method of abstraction: 2004

Data element	Number	Item response rates[1]			
		Total	Self-CADE	Field-CADE	Data-CADE
Total student record respondents	81,810	100.0	100.0	100.0	100.0
Student characteristics					
Date of birth	75, 460	92.2	98.1	97.5	79.9
Gender	75,930	92.8	98.8	98.0	80.0
Marital status	38,400	46.9	57.4	49.0	29.7
Citizenship	69,120	84.5	94.0	90.0	66.3
Veteran status	59,270	72.5	73.8	74.4	68.9
High school completion type	55,840	76.8	82.9	78.0	67.4
Race	63,850	78.1	86.4	78.7	64.7
Hispanic status	63,830	78.0	84.4	64.1	77.5
At least one phone number	76,570	93.6	94.0	94.2	92.3
At least two phone numbers	34,320	42.0	42.9	44.7	38.6
Enrollment					
Type of degree program	77,850	95.2	97.7	92.9	92.5
Student class level	72,900	89.1	92.5	85.4	86.1
Tuition jurisdiction classification	76,880	94.0	99.4	95.8	84.3
Total tuition amount	75,940	92.8	97.2	91.4	86.9
Financial aid[2]					
Any aid received (Y/N)	81,600	99.8	99.6	99.2	99.9
Federal aid received (Y/N)	81,600	99.8	99.6	99.2	99.9
State aid received (Y/N)	81,600	99.8	99.6	99.2	99.9
Undergraduate aid received (Y/N)	81,600	99.8	99.6	99.2	99.9
Graduate aid received (Y/N)	81,600	99.8	99.6	99.2	99.9
Other aid received (Y/N)	81,600	99.8	99.6	99.2	99.9
Total financial aid amount[3]	81,570	99.7	99.6	99.2	99.9
Expected family contribution (EFC) amount	52,220	90.2	98.7	97.8	75.5

[1] Response rate is based on the number of students to whom the item applied.
[2] All financial aid gate items were logically coded for data-CADE students, based on the presence or absence of amounts in the nested items.
[3] Total financial aid amount was computed by summing the amounts entered for each specific aid program. If the financial aid gate item was missing, then the financial aid amount item was also missing.

NOTE: Mode differences were detected for every item presented in this table. All are significant at the 0.05 level.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

4.3 Student Interviewing

4.3.1 Identification of First Time Beginners (FTBs)

NPSAS:04 serves as the base year of a longitudinal study of students beginning their postsecondary education experience during one of the terms of the NPSAS sample year. An FTB student is one who enrolled in postsecondary education for the first time after high school at some time during the NPSAS year (July 1, 2003–June 30, 2004). Also considered "effective FTBs" are those who had previously enrolled, but had not completed a postsecondary course for credit prior to July 1, 2003. Those determined to be FTBs will be followed at periodic intervals

as part of the Beginning Postsecondary Students Longitudinal Study follow-up surveys (BPS:04/06, BPS:04/09), with the data collected during NPSAS:04 serving as the base year for the subsequent longitudinal studies.

NPSAS:04 is the third NPSAS to "spin off" a cohort of beginning students; NPSAS:90 was the first and NPSAS:96 was the second. Based on past experiences, sampling and screening procedures were implemented that were targeted to yield an adequate number of students *that are accurately identified as FTBs* for the BPS:04 cohort. Procedures specific to this purpose were implemented at almost every step of full-scale study operations (e.g., detailed instructions for institutional identification of FTBs when providing enrollment lists; sample selection procedures; wording of CADE items asked specifically about potential FTBs; comprehensive FTB-eligibility questions in the student instrument to make the final FTB determination; and extra locating/interviewing efforts applied to the sample from the student stratum of potential FTBs). FTB sampling rates were based primarily on NPSAS:96 results. The two major challenges in achieving adequate FTB yields are (1) proper identification of a sufficient base from which to obtain FTBs and (2) locating, identifying, and interviewing FTBs from that base in sufficient numbers.

Locating and interviewing potential FTBs is particularly important, since final FTB determination rests on student responses to specific questions.[35] Student records maintained at most postsecondary institutions do not contain all information necessary to make accurate FTB determinations. Insufficiency of institution-level information is quite obvious when considering students who transfer between institutions and may or may not have transfer credits (or other records of such prior education).

Nonetheless, institutions can identify FTBs stochastically; however, instructions to institutions regarding preliminary identification of potential FTBs must also be sufficiently clear and viable that the institution can implement them correctly.[36] Sampling procedures implemented during NPSAS:04 accounted for potential definitional difficulties. As a first screening, institutions were asked to identify potential FTBs according to the following conditions.

Potential FTBs must

- be *undergraduate* students between July 1, 2003 and April 30, 2004;
- have enrolled at the institution *for the first time* between July 1, 2003 and April 30, 2004;
- be classified by the institution as *freshman, or first-year student* at the time of that first enrollment; and
- have no transfer credits from another postsecondary institution.

Based on prior experience, it was anticipated that two types of errors would still exist in lists provided by the schools; specifically, (1) students listed as potential FTBs would not be

[35] A number of questions were contained in the student interview to screen for first-time beginner (FTB) status, including when the student first attended a postsecondary institution, whether the student received any prior postsecondary degrees or certificates, and whether the student completed the first class toward a postsecondary degree or certificate after high school at a postsecondary institution.

[36] Simply asking the institution to identify students who enrolled in the institution for the first time is insufficient, since it can result in identification of undergraduate transfer students as well as first-time enrolling graduate and first-professional students.

actual FTBs (a false positive group) and (2) students not identified as potential FTBs would, in fact, prove to be FTBs (a false negative group). The actual BPS:04 cohort would, thus, consist of those in the potential FTB group *minus* the identified false positives in that group *plus* any false negatives identified in other student strata. Because experience with NPSAS:96 indicated that the false positive rate would exceed (considerably) the false negative rate (Riccobono et al. 1997), the potential FTB stratum was oversampled (see chapter 2). Information to determine FTB status was also collected during CPS matching and record abstraction (CADE).

The student interview FTB screening was accomplished very early in the interview (immediately following NPSAS study eligibility determination).[37] The FTB screening questions were asked of all interviewed undergraduate students so that false positives from the potential FTB stratum could be eliminated from the BPS cohort and so false negatives from the other student strata could be identified and included in the BPS cohort.

The final FTB determination was made based on the student interview. However, there are students who were not interviewed but are potential FTBs based on data obtained from institutional records and/or CPS data. Table 31 provides the results of interview-based FTB determination by initial student classification. Overall, 40 percent of the students interviewed (25,000 students) were determined to be FTBs. Among those initially sampled as potential FTBs based on the list acquisition process, 69 percent were confirmed as FTBs, yielding a 31 percent false positive rate. Among students sampled as "other undergraduates," 25 percent were also determined to be FTBs (false negatives.) The false positive and false negative rates reveal the difficulties that many schools experienced in accurately identifying FTBs.

Table 31. First-time beginner (FTB) determination, by student type: 2004

Sampled student type	Students interviewed[1]	Confirmed FTBs	
		Number	Percent[2]
All students	62,130	24,930	40.1
Total undergraduate	53,590	24,900	46.5
Potential FTB	26,040	18,030	69.2
Other undergraduate	27,540	6,870	24.9
Graduate/first-professional	8,540	30	0.4

[1] Includes study respondents who completed the student interview, since confirmation of FTB eligibility status required contact with the sample members.

[2] Percent is based on the number of students within the row under consideration.

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

4.3.2 Data Collection Evaluations

Help desk

As described in chapter 2, a help desk was available to assist respondents in completing the student interview. Help desk staff were trained to answer any calls received from the help desk hotline, as well as conduct telephone interviews as needed. Help desk staff assisted sample members with questions about the Web instrument and provided technical assistance to sample

[37] First-time beginner (FTB) status was determined at the start of the student interview because many subsequent questions were to be asked only of the actual Beginning Postsecondary Students Longitudinal Study (BPS) cohort.

members who experienced problems while completing the self-administered Web interview. Help desk agents also responded to voice-mail messages left by respondents when the call center was closed. To gain a better understanding of the problems encountered by students attempting to complete the interview, a software program was developed to record each help desk incident that occurred during data collection. For each occurrence, help desk staff confirmed contact information for the sample member, recorded the type of problem, a description of the problem and resolution, incident status (pending or resolved), and the approximate time it took to assist the caller.

Table 32 summarizes help desk incidents encountered during student data collection. Of all calls to the help desk, about 93 percent called the help desk only once, while 6 percent called twice, and 1 percent called three or more times. Of the students who called the help desk, 86 percent completed either a full, Spanish, abbreviated, or partial interview either on their own or with the telephone agent who took their call. The remaining 14 percent did not complete the interview.

Table 32. Help desk incidents, by type: 2004

Type of incident	Total incidents recorded	Percent of total incidents
Total	1,849	100.0
Study ID/password	1,199	61.0
Called in to complete the interview	203	10.3
Questions about the study	195	9.9
Browser settings/computer	171	8.7
Website down/unavailable	45	2.3
Questionnaire content	21	1.1
Program error call-in	14	0.7
Routing/skip problems	1	0.1
Other	118	6.0

NOTE: Details may not sum due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

The majority of the help desk contacts were requests for study ID and/or password (61 percent). Ten percent of calls to the help desk were to complete a telephone interview. Other calls to the help desk regarded general questions about the study (10 percent), problems with browser settings and computer or both (9 percent), and calls to report the website being down or unavailable (2 percent).

Response burden and effort

Time to complete the student interview. The time burden associated with completion of the NPSAS:04 interview was calculated separately for each mode of data collection: self-administered and computer-assisted telephone interview (CATI).

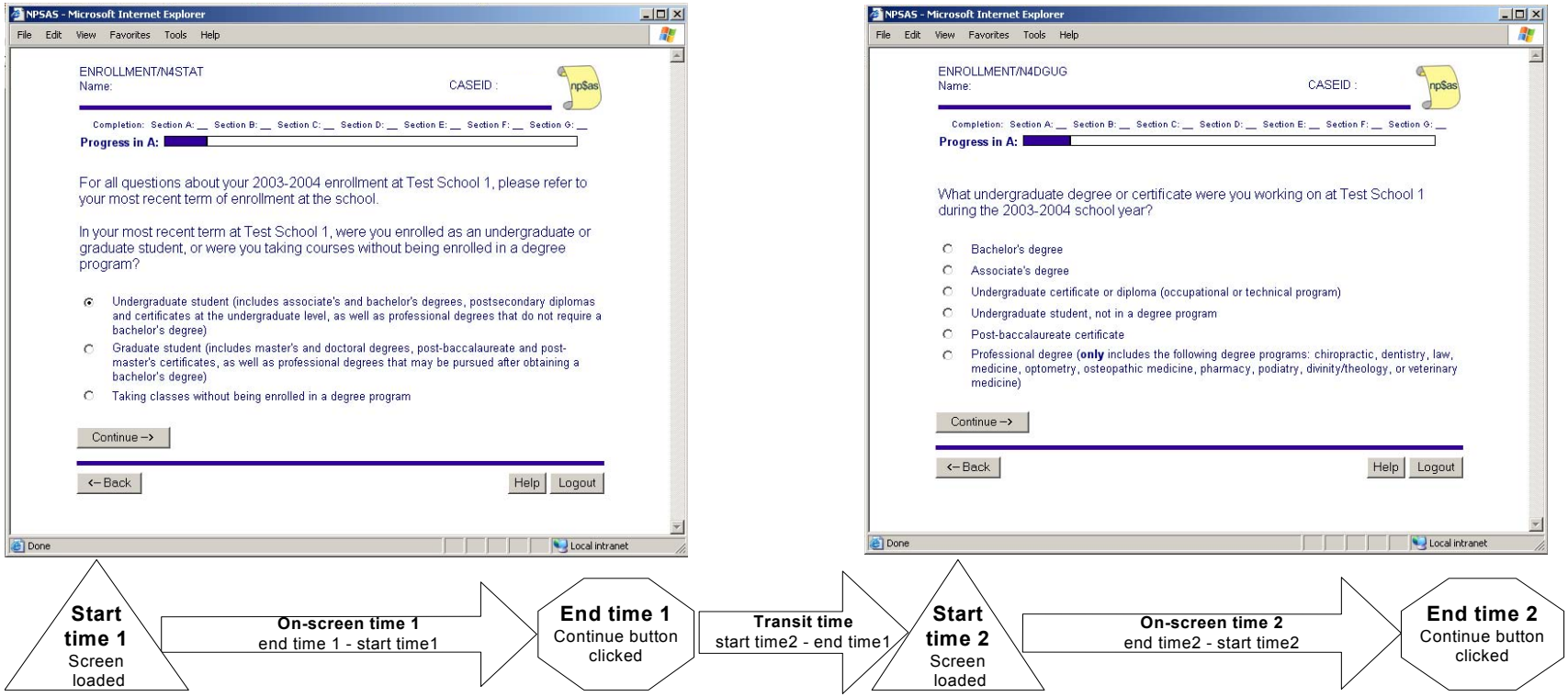
Figure 12 provides a visual representation of how the on-screen and transit times were determined. Two time stamp variables were associated with each interview question. The first, the start timer, was set to the clock time on the respondent's or interviewer's computer at the time that a particular Web page was displayed on the screen. The second time stamp variable, the end timer, was set to the clock time on the respondent's or interviewer's computer at the moment

the respondent or interviewer clicked the "Continue" button to submit the answers from that page.

From the two time stamp variables, an on-screen time and transit time were calculated. The on-screen time was calculated by subtracting the start time from the end time for each Web page that the respondent received. The transit time was calculated by subtracting the end time of the preceding page from the start time of the current page; it includes the time required for the previous page's data to be transmitted to the server, for the server to store the data and assemble and serve the current page, and for the current page to be transmitted to and loaded on the respondent's or interviewer's computer.

A total on-screen time was then calculated for all respondents by summing the on-screen times for each Web page that the respondent received. For each respondent, a total transit time was calculated by summing all the transit times. The total on-screen and total transit times were then summed to determine the total instrument time.

Figure 12. Visual representation of on-screen and transit times: 2004



Total On-screen time = On-screen time 1 + On-screen time 2 +…+ On-screen time N
Total Transit time = Transit time 1 + Transit time 2 +…+ Transit time N
Total Instrument time = Total On-screen time + Total transit time

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Table 33 presents the average times for the full interview overall and by student type. The average time to complete the entire interview was about 27 minutes. The interview was longest for FTBs (31 minutes,) largely because they received additional questions not applicable to other students. Total interview time took about 25 minutes for other undergraduates and 20 minutes for graduate and first-professional students ($t = 57.59$, $p < .0001$).

Table 33. Average time to complete full-scale student interview, by student type: 2004

Interview section	Number of cases	Average time
All students	52,560	26.5
FTB student	20,770	31.1
Other undergraduate	24,320	24.6
Graduate/first-professional student	7,470	20.0

† Not applicable.

NOTE: Detail may not sum to totals because of rounding. FTB = first-time beginner. Outliers were excluded from this analysis. Outliers were identified separately for each section and for the total interview; therefore, individual section times do not sum to the total interview times. An outlier was defined as any case whose completion time exceeded two standard deviations above or below the average time for a given section. Interview times are presented only for completed interviews (partial interviews were excluded).

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Table 34. Average time to complete full-scale student interview, by interview section and student type: 2004

	All respondents		FTB student		Other undergraduate		Graduate/first-professional student	
Interview section	Number of cases	Average time	Number of cases	Average time	Number of cases	Average time	Number of cases	Average time
Section A—Enrollment	55,790	7.8	22,030	7.8	25,850	8.2	7,910	6.6
Section B—Financial aid	55,950	3.9	22,070	4.0	25,940	3.9	7,950	3.9
Section C—Employment	56,070	4.7	22,130	4.7	26,020	5.0	7,920	3.2
Section D—Education experiences	56,220	2.7	22,150	5.4	26,120	1.1	7,960	0.5
Section E—Background	55,870	4.7	22,040	4.9	25,880	4.8	7,950	4.1
Section F—Locating	22,080	3.8	22,080	3.8	†	†	†	†

† Not applicable.

NOTE: Detail may not sum to totals because of rounding. FTB = first-time beginner. Outliers were excluded from this analysis. Outliers were identified separately for each section and for the total interview; therefore, individual section times do not sum to the total interview times. An outlier was defined as any case whose completion time exceeded two standard deviations above or below the average time for a given section. Interview times are presented only for completed interviews (partial interviews were excluded).

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Table 34 presents the average times for each section overall and by student type. The first section on enrollment collected key information necessary for eligibility determination and FTB identification, as well as information about enrollment, degree program, and field of study[38]. Much of the critical information needed to assess student status and other characteristics necessary for routing to the appropriate questions in the remainder of the interview was collected in this initial section. This was the longest section, taking just under 8 minutes to complete.

The second section focused on financial aid. It contained items about school-related jobs such as work-study and assistantships, as well as questions about other forms of financial aid such as grants, loans, and scholarships. Additional items asked about parental support and the

[38] See the student interview facsimile in Appendix E for more detail about the content of the interview sections.

use of educational tax credits. Overall, respondents took an average of 4 minutes to complete this section.

The employment section collected information on jobs held while enrolled, balancing school and work, and assets and debts. This section took approximately 5 minutes to complete.

The section on education experiences contained a few items applicable to all respondents, such as the items about distance education. However, many items were administered only to FTBs, such as those focusing on undergraduate experiences, transfers, and factors related to choice of postsecondary institution. This section averaged about 3 minutes overall, but took 5 minutes for FTBs, 1 minute for other undergraduates, and less than 1 minute for graduate and first-professional students.

The background section focused on basic demographics about the students and their families. Citizenship status, community service, and education-related disabilities were also topics of interest in the background section. Overall, the average time to complete this section was about 5 minutes.

The final section applied only to FTBs for the purpose of collecting locating information for future follow-up studies with this cohort. FTBs took an average of 4 minutes to complete this section.

Interview times were also evaluated by mode of administration. Table 35 shows the total interview time. The difference in total interview completion by mode was small but significant; approximately 26 minutes for self-administered respondents and 27 minutes for interviewer-administered respondents ($t = 8.92$, $p < .0001$).

Table 35. Average time to complete full-scale student interview, by interview section and mode of administration: 2004

	All respondents		Self-administered respondents		Interviewer-administered respondents	
Interview section	Number of cases	Average time	Number of cases	Average time	Number of cases	Average time
Section A—Enrollment	55,790	7.8	25,280	7.9	30,520	7.7
Section B—Financial aid	55,950	3.9	25,330	4.3	30,630	3.7
Section C—Employment	56,070	4.7	25,490	4.4	30,580	4.8
Section D—Education experiences	56,220	2.7	25,820	2.4	30,400	3.0
Section E—Background	55,870	4.7	25,190	5.0	30,680	4.5
Section F—Locating	22,080	3.8	8,980	3.8	13,100	3.7

NOTE: Detail may not sum to totals because of rounding. Outliers were excluded from this analysis. Outliers were identified separately for each section and for the total interview; therefore, individual section times do not sum to the total interview times. An outlier was defined as any case whose completion time exceeded two standard deviations above or below the average time for a given section. Interview times are presented only for completed interviews (partial interviews were excluded).

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Table 36 presents the average time on-screen and in transit by response mode. Average transit times were twice as long for self-administered respondents than for interviewer-administered respondents (6 minutes and 3 minutes, respectively; $t = 90.03$, $p < .0001$). On-screen times were significantly less for self-administered respondents than for interviewer-administered respondents (20 minutes and 24 minutes, respectively; $t = -53.95$, $p < .0001$.)

It is likely that interviewer-administered respondents took slightly longer to complete the interview sections because respondents and interviewers were engaged in a conversation, and respondents had to wait for interviewers to read the entire question and response options (depending on the nature of the screen and the interviewer instructions[39]). Self-administered respondents, however, could read and respond to interview questions more quickly because they were able to read the entire screen at once.

Table 36. Average on-screen and transit time, by response mode: 2004

Response mode	Number of cases	Average total interview time	Average on-screen time	Average transit time
All respondents	52,560	26.5	22.3	4.2
Self-administered respondents	22,100	26.1	20.4	5.8
Interviewer-administered respondents	30,460	26.8	23.7	3.1

NOTE: Detail may not sum to totals because of rounding. Outliers were excluded from this analysis. Outliers were identified separately for each section and for the total interview; therefore, individual section times do not sum to the total interview times. An outlier was defined as any case whose completion time exceeded two standard deviations above or below the average time for a given section. Interview times are presented only for completed interviews (partial interviews were excluded).

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

At the end of the survey, a short debriefing section asked questions about users' experiences in completing the Web survey. As part of the debriefing section, self-administered respondents were asked which type of internet connection they used to access the survey. Table 37 presents the average total interview times and transit times by type of internet connection.

Among self-administered respondents, about 6 percent completed the interview through a dial-up modem, and about 41 percent completed with a fast connection (i.e., cable modem, DSL, ISDN, LAN). Total interview time for dial-up modem connections was nearly 35 minutes, compared to 24 minutes for those using a fast connection ($t = –49.24$, $p < .0001$). This large variation can be attributed to transit times, which were also much higher for the dial-up connection versus the fast connections (13 minutes compared with 5 minutes; $t = –85.82$, $p < .0001$).

[39] To minimize mode differences and ensure that all respondents were exposed to the same information, interviewer instructions were included on every form of the questionnaire for computer-assisted telephone interviews (CATIs). These instructions indicated to interviewers how to handle response options (e.g., whether the response options should be read aloud or not).

Table 37. Average time to complete self-administered student interview, by internet connection type: 2004

Internet connection type	Average total interview time	Transit time	Percent of time in transit
Dial-up modem	34.6	13.3	38.5
Fast connection	24.4	4.6	18.7
Cable modem	24.7	4.4	18.0
Digital subscriber line (DSL)	24.8	4.8	19.3
Integrated services digital network (ISDN)	23.4	4.9	20.9
Corporate local area network (LAN; T1 or T3)	23.7	4.5	19.0
Do not know connection type	27.7	6.1	22.2
Other	27.8	6.6	23.6

NOTE: At the end of the interview, a debriefing section was included that asked questions about self-administered respondents experiences in completing the Web survey. Data presented here are based on the self-administered respondents who answered the debriefing questions. Fast connection is the average interview time of respondents with a Cable Modem, Digital Subscriber Line, ISDN, or Corporate LAN. Average total time is sum of on-screen and transit times.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Number of calls

A total of 94,503 telephone interviewer hours (exclusive of training, supervision, monitoring, and administration) were expended to obtain completed interviews from 62,130 sample members. Since the time to administer the interview was, on average, under 30 minutes, the large majority of interviewer time was spent on other case-related activities. A small percentage of this time was required to bring up a case, review its history, and close the case (with appropriate reschedule, comment, and disposition entry) when completed. The bulk of the time, however, was devoted to locating and contacting sample members.

Table 38 shows the average number of calls per case, by interview status and administration mode. The overall average was about 14 calls per case. Among all completed cases, an average of 10 call attempts was required, while the average for nonrespondents was about 21 calls.

Table 38. Average calls per case, by interview status: 2004

Interview status	Number of cases	Number of calls	Mean calls per case
Total	101,010	1,394,948	13.8
Interviewed	62,220	598,556	9.6
Not interviewed	38,800	796,564	20.5
By mode			
Self administered—no telephone follow-up	17,040	†	†
Self administered—with telephone follow-up	11,670	195,589	16.8
Interviewer administered	33,510	402,790	12.0

† Not applicable.

NOTE: Detail may not sum to totals because of rounding. Excludes 8,200 cases determined to be ineligible for the study.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

The average call count varied by mode of data collection. Of the 62,220 completed cases, approximately 27 percent were completed via self-administration and required no telephone prompting. However, an average of 17 calls was made to the remaining 11,670 self-administered cases to encourage interview completion. Finally, approximately one-half of the completions (54 percent) were obtained by a telephone interviewer and required an average of 12 call attempts.

As seen in table 39, the number of calls per case varied by type of students and type of institution. On average, potential FTB students and other types of undergraduates required more calls (14 calls) than graduate and first-professional students (11 calls) ($F = 314.6$, $p < 0.001$). Additionally, those from less than 2-year institutions and 2-year institutions required more calls on average (16 calls and 14 calls, respectively) than those from either 4-year doctorate-granting or 4-year non-doctorate-granting institutions (13 calls and 14 calls, respectively) ($F = 111.5$, $p < 0.001$).

Table 39. Average calls per case, by institutional characteristics and student type: 2004

Institutional characteristics and student type[1]	Number of cases	Number of calls	Mean calls per case
Total	101,010	1,394,948	13.8
Institutional level			
Less than 2-year	10,330	166,830	16.2
2 Year	37,750	529,633	14.0
4-year non-doctorate-granting	20,630	283,250	13.7
4-year doctorate-granting	32,310	415,507	12.9
Institutional control			
Public	65,540	889,378	13.6
Private not-for-profit	21,660	289,161	13.4
Private for-profit	13,820	216,421	15.7
Type of institution			
Public less than 2-year	2,150	33,841	15.7
Public 2-year	32,540	453,282	13.9
Public 4-year non-doctorate-granting	8,890	120,548	13.6
Public 4-year doctorate-granting	21,960	281,966	12.8
Private not-for-profit 2-year or less	2,730	37,046	13.6
Private not-for-profit 4-year non-doctorate-granting	8,880	122,189	13.8
Private not-for-profit 4-year doctorate-granting	10,050	129,746	12.9
Private for-profit less-than-2-year	7,550	124,802	16.5
Private for-profit 2-year or more	6,260	91,396	14.6
Student type			
Total undergraduate	89,460	1,275,252	14.3
FTB student	39,440	558,470	14.2
Other undergraduate	50,020	717,787	14.4
Graduate/first-professional	11,560	118,837	10.3

[1] Both institutional and student classifications were verified to correct classification errors on the sampling frame.

NOTE: Detail may not sum to total due to rounding. Excludes 8,200 cases determined to be ineligible for the study. FTB = first-time beginner.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

4.3.3 Instrument Usability

Coding

The NPSAS:04 student interview obtained students' field of study by first collecting a verbatim string and then providing a list of options from which the appropriate category could be selected. To assess the accuracy of coding procedures, a random sample of 10 percent was selected from all strings provided. Expert coders evaluated the verbatim strings for completeness and for the appropriateness of the assigned codes, determining whether a different code should have been assigned or if a string was too vague to code.

Table 40 provides the results of the coding analyses. Of all the strings analyzed, 79 percent were coded correctly. The coding results for major field of study were similar between modes of data collection, indicating that expert coders agreed with self-administered respondent coding at about the same rate as they agreed with interviewer-administered interview coding ($\chi^2 = 0.79$, $p > 0.05$). The quality of the text strings was high, with only 2 percent of text strings too vague to be coded.

Table 40. Summary of coding results for major field of study, by respondent type: 2004

Respondent type	Coding attempts sampled	Percent original code correct	Percent text string too vague to code
Total	4,598	78.8	2.0
Self-administered	2,361	79.3	1.9
Interviewer-administered	2,237	78.3	2.1

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Help text usage

Each Web screen in the NPSAS:04 instrument was equipped with help text to aid respondents with general and screen-specific instrument inquires. The instrument provided general help text which outlined basic information on internet browsers and response types (i.e., how to use a check box, drop-down, or radio button). Each help text screen provided a toll-free number to the NPSAS:04 help desk for further questions. The screen-specific help text defined instrument vocabulary, instructed respondents on how to enter responses, and explained the type of information requested for each form.

Counters placed within the instrument calculated the number of times help text for each screen was accessed. These were analyzed overall and by administration mode to determine which screens may have been problematic for users.

The screen-level rate of help text access was below 2 percent for most of the screens in the NPSAS:04 interview. Help text access rates were analyzed overall and by administration mode. Across all interview forms, cases completed with an interviewer accessed help text more often than did self-administered cases (1.3 percent compared with 0.3 percent, respectively; $t = 6.43$, $p < 0.0001$). Table 41 presents the interview screens[40] for which help text was accessed at a rate of 5 percent or more, based on the number of cases to whom the form was administered. Differences by administration mode are all significantly different ($p < 0.0001$) with interviewer-

[40] See appendix E for a facsimile of the student interview questionnaire.

administered cases accessing help text more frequently than self-administered respondents. It should be noted that interviewers were trained to use help text, whereas self-administered respondents may have forgotten it was available.

Table 41. Rates of help text usage for items accessed by 5 percent or more of respondents, by interview screen and administration mode: 2004

Screen name	Description	Number administered to	Number of help text accesses	Percent		
				All modes	Self-administered	Interviewer-administered
N4ASSOC	Type of associate's degree	12,770	1,300	10.2	2.8	14.2
N4CLSLV	Class level for non-degree students	6,990	570	8.2	1.9	11.8
N4SCHJOB	School-related job	52,490	3,540	6.7	1.9	10.4
N4VOCREC	Received vocational rehabilitation services	5,900	390	6.6	1.1	10.7
N4GRAID	Graduate assistantships	9,290	480	5.2	1.9	10.6
N4EMPTYP	Type of employer	39,680	2,000	5.0	0.5	8.7
N4OTAID	Other aid received	13,920	660	4.7	1.3	9.3
N4CMPCLS	Completed postsecondary class before 7/1/2003	21,400	1,010	4.7	1.9	6.4

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

N4ASSOC had the highest rate of help text accesses. Among students who were administered this form, 10 percent used help text. This form was asked of students who reported that they were working on an associate's degree. It was a follow-up question to differentiate between Associate of Arts (AA) and Associate of Science (AS) degrees. Self-administered respondents (3 percent) were less likely than interviewer-administered respondents (14 percent) to seek help text for this form ($Z = -20.50$, $p < 0.0001$). This result is likely due to the way the question was asked; "What type of associate's degree were you working on at [NPSAS institution]?" While self-administered respondents could read the response options and immediately understand the intent of the question, those who completed a telephone interview did not receive the same visual cues. This item will be revised in future studies to minimize the mode difference observed here.

N4CLSLV asked non-degree students to classify themselves as primarily undergraduate, graduate, or an equal mix of both. The overall help text rate was about 8 percent but was primarily used by CATI respondents (12 percent compared with 2 percent for self-administered respondents; $Z = -14.54$, $p < 0.0001$).

N4SCHJOB asked respondents if they participated in a work study or paid assistantship through their institution. The help text usage rate for this screen was about 7 percent. CATI yielded the most help text hits for N4SCHJOB with an average of 10 percent, compared to 2 percent for self-administered respondents ($Z = -38.33$, $p < 0.0001$).

Respondents who had indicated having some type of disability were asked to report whether or not they had received Vocational Rehabilitation in N4VOCREC. Valid response options for this screen consisted of only yes or no answers. The help text rate was about 7 percent overall.

N4GRAID was a form that contained several check-box items that asked graduate students about graduate assistantships and aid amounts. It asked about teaching and research

assistantships, as well as other less common types (traineeships). Help text provided definitions of each type of graduate aid listed. It is likely that respondents were seeking the definitions for the less common types of aid on this form. Of all graduate students who were administered this form, 5 percent used help text.

The help text rate for N4EMPTYP was about 5 percent. N4EMPTYP asked respondents to categorize their employer type among six options (the NPSAS institution, a for-profit company, nonprofit organization, military, self employed, or local, state, or federal government). Response options were read to interviewer-administered respondents to ensure that they would know what the choices were, as did self-administered respondents. However, the help text rate was still 5 percent overall (1 percent for self-administered and 9 percent for interviewer-administered respondents; $Z = -37.07$, $p < 0.0001$).

N4OTAID was a screen that asked respondents about alternative sources of financial aid not administered through institutional financial aid offices. Items focused on employer aid (both the student's and parents' employers), aid from private organizations, and veteran's benefits. This is information that has traditionally been very hard to collect from students because many do not know, which likely explains the high rate of help text access (5 percent).

N4CMPCLS was a critical item used in the final determination of FTB eligibility status. It was asked of any undergraduate who appeared to be an FTB but who had possibly enrolled in postsecondary education prior to the beginning of the NPSAS year. It asked whether students had ever completed a postsecondary course for credit prior to enrolling at the NPSAS institution. Among students who were asked this question, 5 percent used the help text as a reference prior to providing an answer.

4.3.4 Item Nonresponse

Critical item conversion

As noted earlier, NPSAS:04 is the first cycle to provide the option for self-administration of the student instrument. To minimize item-level nonresponse for certain key items, conversion text was displayed to emphasize the confidential nature of the study and reiterate the importance of individual responses. These items focused on enrollment status and dates, the employment history of the respondent, and parent income.

If a respondent did not answer one of the six items (i.e., left the item blank and hit the continue button), the item screen was reloaded with additional text emphasizing the importance of the item. For some items, a "don't know" option was added to determine if the initial nonresponse was for that reason. The intent was to encourage respondents to provide an answer to the item and to discern the reason for leaving the item blank originally (e.g., refusal or did not know the answer).

Overall, conversion text was moderately successful in converting blank responses either to a valid response or to a don't know response. Results are presented in table 42. The percent of initially blank responses subsequently converted to a valid response ranged from 21 percent for parents' income to 87 percent for student status. There were no differences between self-administered and telephone interviews in rates of conversion, with one exception. Critical item text conversion was more successful for self-administered interviews than telephone interviews for obtaining valid responses to the number of jobs held ($t = 42.80$, $p < 0.05$).

Table 42. Conversion rates for critical items: 2004

Variable	Cases viewing conversion text	Percent converted by subsequently providing a valid response	Percent converted by subsequently providing a "don't know" response	Total percent converted
Student status at NPSAS	180	87.4	†	87.4
NPSAS enrollment by month[1]	400	57.3	34.3	91.6
Date first began NPSAS[1]	390	65.1	0.0	65.1
Date first attended school	360	47.0	†	47.0
Number of jobs during NPSAS year	440	65.8	†	65.8
Parents' income 2002[1]	7,260	21.1	69.7	90.8

† Not applicable.
[1] For these items, a "don't know" response option was added when the screen reloaded, in addition to text emphasizing the importance of the item.
NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Three of the items that presented conversion text also displayed a "don't know" response option when the screen was shown for a second time: NPSAS enrollment by month, date of first attendance at the NPSAS school, and parents' income.

For NPSAS enrollment by month, 34 percent of the cases who initially provided no response reported "don't know" when the conversion text was displayed, resulting in a total conversion of 92 percent of all initially blank responses to either a valid response or a "don't know."

The "don't know" option was selected by 70 percent of all respondents who did not provide an initial response to the question about parents' income, yielding a total conversion rate of 91 percent to either a valid or don't know response.

While the "don't know" option was presented when the question about date of first enrollment at the NPSAS school, it was not selected by any respondents who saw the conversion text. This result is likely due to the format of the response options. Respondents were instructed to select their answers from two drop-boxes: one for month and one for year. The "don't know" option was embedded within the drop-boxes, and it is likely that respondents did not see the new options when the screen was re-displayed. In future studies, this format will be revised so that the "don't know" option is more visible to respondents.

Item-level nonresponse

All respondents to the student interview were provided the option to decline to answer any item. In previous rounds of the NPSAS survey, interviewers were provided with one of two options for this purpose: "don't know" and "refused." In NPSAS:04, the don't know response was only available for key items and provided only as a follow-up option when the screen was initially left blank. Respondents may have given a don't know response for a number of reasons. The most obvious is that the answer is truly unknown or in some way inappropriate for the respondent. Don't know responses may also be evoked when the question wording is not understood by the respondent or when the respondent hesitates to provide a "best guess" response. If respondents failed to give a valid answer or to respond "don't know," their response was considered "blank." There was no explicit "refusal" option in NPSAS:04. This section

presents the results of an analysis of missing data among student interview respondents to better understand which items may be sensitive or difficult to answer.[41]

Item nonresponse rates were calculated for items asked of at least 100 respondents. Item nonresponse rates in the NPSAS:04 interview were low, with 24 items of approximately 210 items containing over 10 percent missing data. These items are shown in table 43 and grouped by interview section. Most nonresponse resulted from respondents leaving the item blank. Five of these 24 items were missing values due to respondents reporting that they did not know the answer.

Table 43. NPSAS:04 interview item nonresponse for items with more than 10 percent missing

Interview section	Variable name	Description	Number asked	Percent "don't know"	Percent blank	Total percent nonresponse[1]
Section A: Eligibility and Enrollment	N4PRBA	Earned bachelor's while a first-professional student	1,510	†	13.0	13.0
	N4MAJ2A	Major-secondary string	1,420	†	10.8	10.8
	N4MAJ2B	Major-secondary category	1,420	†	9.8	9.8
	N4LT30	Age: less than 30	390	†	17.7	17.7
	N4SCH2	School 2 name	540	0.0	13.9	13.9
	N4CT2	School 2 city	540	0.0	11.3	11.3
	N4LEVL2	School 2 level	540	0.2	15.0	15.2
	N4CTRL2	School 2 control	540	0.0	15.4	15.4
Section B: Financial Aid	N4TASSM	Teaching assistantship amount	1,240	†	9.6	9.6
	N4RASSM	Research assistantship amount	1,240	†	9.6	9.6
	N4TRNSM	Traineeship amount	130	†	21.3	21.3
	N4GASSM	Other graduate assistantship amount	340	†	13.1	13.1
	N4STAMT	State grant/scholarship amount	8,310	†	13.0	13.0
	N4AMNEMP	Amount of employer aid	3,960	†	11.2	11.2
	N4AMNVET	Amount of veteran's benefits	1,610	†	18.5	18.5
	N4AMNPMP	Amount of parents' employer aid	1,080	†	16.6	16.6
Section C: Expenses	N4HOPE	Claim Federal Hope scholarship	59,220	31.9	5.4	37.3
	N4DEDUCT	Claim tuition tax deduction	59,250	33.6	3.7	37.3
	N4LFLNG	Claim lifetime learning tax credit	59,070	33.0	4.9	37.8
	N4PARNC	Parents income in 2003	40,210	12.6	1.7	14.3
Section E: Background	N4TRIBE	State/federally recognized tribe	1,380	†	13.1	13.1
	N4RACES	Race: other specify	6,870	†	17.6	17.6
	N4SERCS	Service: other specify	440	†	20.3	20.3
	N4NEEDS	Needs: other specify	510	†	32.9	32.9

† Not applicable.

[1] Item nonresponse rates were calculated based on the number of student interview respondents for whom the item was applicable and asked.

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

The item with the highest rate of nonresponse in the student eligibility and enrollment section pertained to date of birth. Respondents who did not provide a date of birth were asked to provide a categorical age range (N4LT30). Of the respondents who did not provide a date of birth, about 18 percent also failed to provide a categorical age range. Students in first-professional programs were asked whether they had completed a baccalaureate degree in order to determine student status (N4PRBA). About 13 percent of students to whom this item was

[41] See chapter 6 and appendix K for analyses of nonresponse bias among all study respondents.

administered failed to provide a response. Two related items collected information about the second major (N4MAJ2A, N4MAJ2B; a verbatim string and categorical major code) for students working on a double major, which was missing for approximately 10 percent of cases. For students who attended other institutions in addition to the NPSAS school between July 1, 2003, and June 30, 2004, information was collected on the other institutions attended (N4LEVL2, N4CTRL2). The items pertaining to the level and control of the other institution were both collected for schools not codeable within the online IPEDS coding system. These items were missing for about 15 percent of respondents to whom these items applied.

The financial aid section contained several forms that collected information about different types of financial aid received. In one series of items, graduate students were asked whether they had different types of assistantships or a traineeship, and then those who indicated having such aid were asked to provide a dollar amount (N4TASSM, N4RASSM, N4TRNSM, N4GASSM). About 10 percent of cases with either a graduate teaching or research assistantship did not provide a dollar amount. About 13 percent of those reporting another graduate assistantship did not provide a dollar amount, and 21 percent of students with a traineeship also left the dollar amount blank. In another series of items regarding financial aid, students were asked whether they had received certain types of aid not administered through the institutional financial aid office, including employer aid and veteran's benefits (N4AMNEMP, N4AMNVET, N4AMNPMP). Rates of missing data ranged from 11 to 19 percent for the dollar amount items associated with these types of financial aid. Finally, about 13 percent of students who reported receiving a state grant or scholarship did not provide a dollar amount (N4STAMT).

Items with the highest rates of nonresponse were from the section on expenses and pertained to tax deductions. The following three items were collected on one screen. Students were asked "whether or not they claimed a lifetime learning tax credit" (N4LFLNG), and only 62 percent provided a valid response. Additionally, 37 percent of respondents had missing information on "whether or not they claimed a tax deduction for receipt of the Federal Hope scholarship" (N4HOPE) and "whether or not they claimed a tax deduction for tuition" (N4DEDUCT). The majority of nonresponse for these items was "don't know" rather than "blank." A substantial portion of respondents failed to provide information about their parents' financial situation. Despite the use of conversion text (described in the previous section), about 14 percent of respondents contain missing data on their parents' income (N4PARNC). This is mostly because they do not know this information: about 13 percent do not know their parents' income, while 2 percent left the item blank.

Among the student background variables, items with the highest rates of nonresponse were those asked to respondents who reported having a disability. Of these, 33 percent had missing information on "other disability-related services and accommodations needed to assist with schooling that was not received" (N4NEEDS), and 20 percent had missing information on "other disability services or accommodations received to assist with schooling in the last 12 months" (N4SERCS).

In the telephone interview, attempts to convert item-level nonresponse are from a trained interviewer, while in the self-administered interview prompts to obtain answers for nonresponse are read by the respondent from a computer screen. It is important to understand which items, if any, are difficult for self-administered respondents to understand because they do not have the additional assistance of a trained interviewer while completing the interview. Therefore, in

addition to the overall analysis, item-level nonresponse was analyzed by administration mode. Items with 10 percent or more missing data in either mode are presented in table 44.

Table 44. NPSAS:04 interview item nonresponse for items with more than 10 percent "don't know," by mode of administration

Interview section	Variable name	Variable label	Percent missing in self-administered interview	Percent missing in interviewer-administered interview
Section A: Eligibility and Enrollment	N4PRBA	Earned bachelor's while a first-professional student	6.3	24.8*
	N4MAJ2A	Major-secondary string	8.1	17.2*
	N4MAJ2B	Major-secondary category	6.6	17.2*
	N4GPAEST	Estimate of grade-point average (GPA)	13.4	6.6*
	N4CMPDGN	Completed requirements for degree	1.9	13.1*
	N4LT30	Age: less than 30	17.8	17.7
	N4SCH2	School 2 name	18.5	8.3*
	N4CT2	School 2 city	13.5	8.7
	N4ST2	School 2 state	10.1	6.6
	N4LEVL2	School 2 level	19.9	9.5*
	N4CTRL2	School 2 control	20.5	9.1*
Section B: Financial Aid	N4TASSM	Teaching assistantship amount	9.7	9.5
	N4RASSM	Research assistantship amount	9.5	9.8
	N4GASSM	Other graduate assistantship amount	14.5	9.8
	N4STAMT	State grant/scholarship amount	8.4	17.9*
	N4INAMT	School grant/scholarship amount	4.8	11.3*
	N4AMNEMP	Amount of employer aid	9.6	12.7*
	N4AMNVET	Amount of veteran's benefits	19.1	18.0
	N4AMNPMP	Amount of parents' employer aid	10.1	21.3*
Section C: Expenses	N4DEP03	Claimed as a dependent	11.3	7.4*
	N4HOPE	Claim Federal Hope scholarship	48.7	27.1*
	N4DEDUCT	Claim tuition tax deduction	49.1	26.7*
	N4LFLNG	Claim lifetime learning tax credit	49.6	27.2*
	N4INCSP	Spouse's earnings in 2003	2.3	13.0*
	N4PARNC	Parents' income in 2003	5.6	21.4*
Section E: Background	N4TRIBE	State/federally recognized tribe	6.1	17.9*
	N4DADED	Father's education	6.9	10.9*
	N4RACES	Race: other specify	20.4	16.3*
	N4SERCS	Service: other specify	15.6	23.0
	N4NEEDS	Needs: other specify	47.0	26.0*

* Indicates a significant difference at the 0.05 level.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Twenty items had rates of nonresponse higher than 10 percent among self-administered respondents. Of these, six were unique to those completing the survey online. Four of these items (N4ST2, N4CTRL2, N4SCH2, and N4CT2) were administered to respondents who attended another school in addition to their NPSAS school during the 2003–04 school year. It is possible that respondents were unsure whether and/or how to provide information about multiple postsecondary attendance without the assistance of a trained interviewer. The other two items were self-estimated grade-point average (GPA, N4GPAEST) and whether or not they were claimed as a dependent on their 2003 taxes (N4DEP03).

Twenty-four items had rates of nonresponse higher than 10 percent from respondents who completed a telephone interview. Of these, 10 were unique to CATI respondents. Most of

these were items that inquired about information that could be deemed sensitive, such as personal information and family finances. For example, two were about grants/scholarships (N4STAMT; N4INAMT), two were about income (N4INCSP; N4PARNC), and one was about enrollment in a state or federally recognized tribe (N4TRIBE). It might be the case that respondents felt uncomfortable providing this information to an interviewer.

To discern if there were systematic differences in item nonresponse between interviewer- and self-administered interviews, all items administered to at least 100 respondents and that had at least 10 percent total missing in either self-administration or CATI administration mode were analyzed. The variables meeting this criterion are shown in table 43. For 12 of the 30 items, telephone interviews were more likely than self-administered interviews to have missing information ($p < 0.05$). For another 10 items, self-administered interviews were more likely than telephone interviews to have missing information ($p < 0.05$). There were no significant differences between telephone and Web interviews on 8 items. Items pertaining to sensitive information such as family finances tend to be missing in telephone interviews while items that might require further explanation such as multiple institutional attendance and tax deductions tend to be missing in self-administered interviews.

4.3.5 CATI Monitoring and Quality Assurance

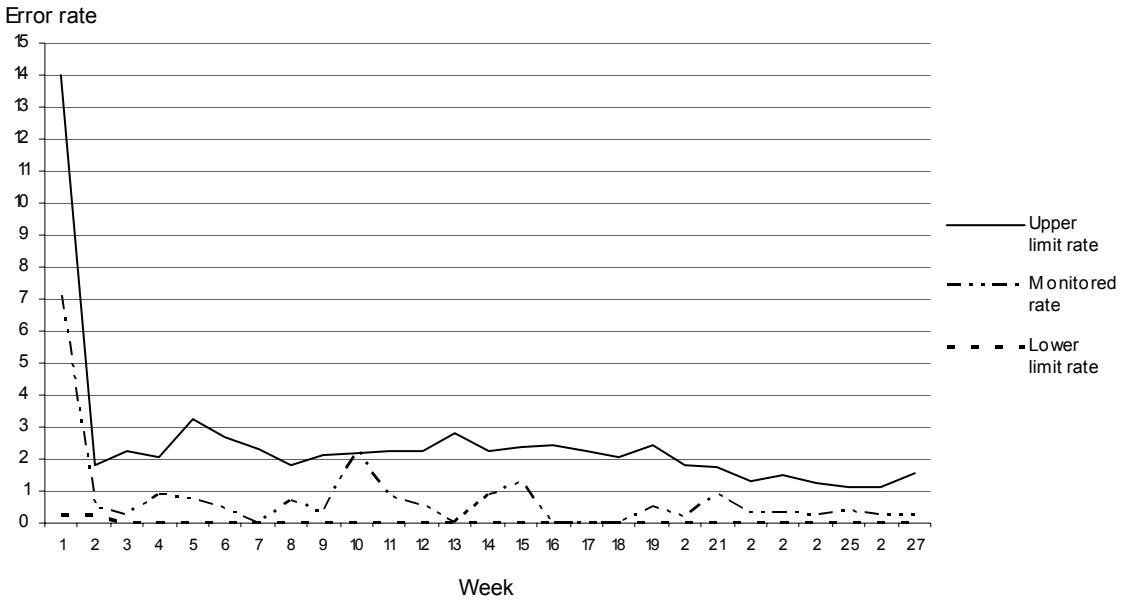
Regular monitoring of telephone interviews leads to better interviewing and data quality as well as improvements in data collection costs and in the efficiency of the telephone facilities. To ensure that sufficient monitoring occurred for the full-scale NPSAS:04, monitoring sessions were conducted during day, evening, and weekend shifts. Monitors listened to and simultaneously viewed the progress of interviews using remote monitoring telephone and computer equipment. Monitors listened to up to 20 questions during an ongoing interview and, for each question, evaluated two aspects of interviewer performance: (1) correct delivery of questions (error in question delivery) and (2) accurate keying of the response (error in data entry).

Measures of question delivery and data entry were developed and daily, weekly, and cumulative reports were produced. Monitoring took place throughout data collection, with a total of 14,775 items monitored. During the initial weeks of data collection, the number of observations was lower because telephone interviews were slow to start. Likewise, monitoring efforts were scaled back during the final weeks of data collection due to lighter caseloads. Among the 14,775 items observed, 77 delivery errors and 25 data entry errors were observed. Error rates in delivery and data entry, by week of data collection are shown in figures 13 and 14, respectively. Overall error rates were low (typically below 2 percent) and within control limits.[42] The peaks in error rates can be attributed to the assignment of new monitors who were learning how to monitor and count errors, and new interviewers who were becoming familiar with the student instrument.[43]

[42] The upper and lower control limits were defined by three times the standard error of the proportion of errors to the number of questions observed for the period (+3 times the standard error for the upper limit; -3 times the standard error for the lower limit). Peaks in control limits can be attributed to variation in the number of observations across weeks.

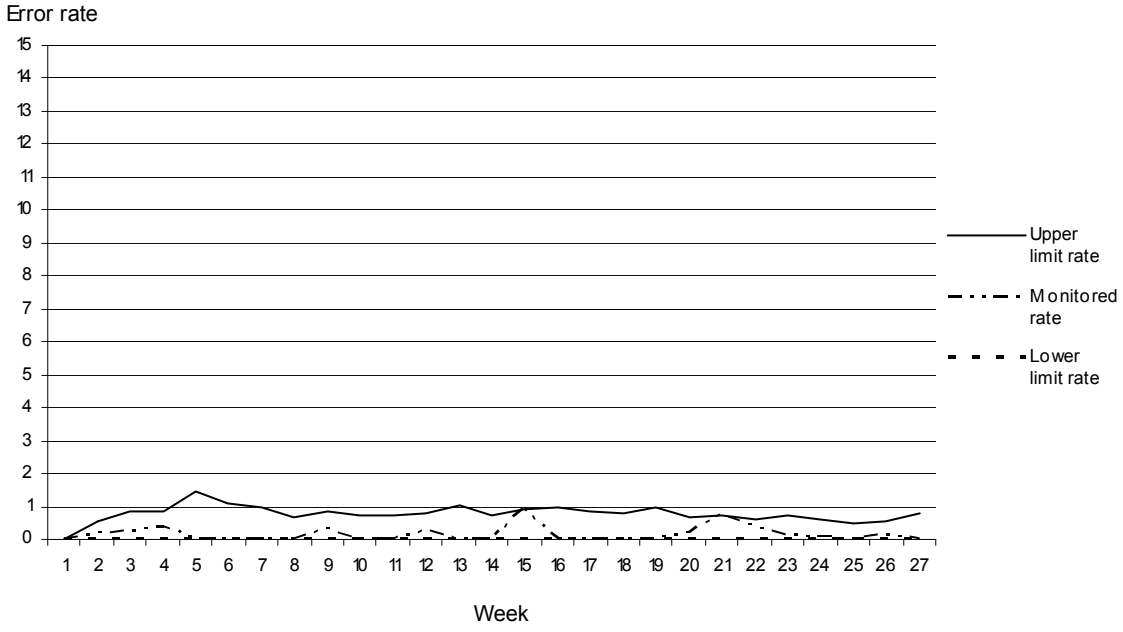
[43] The number of student interviews to be completed required a large interviewing staff. There were several training sessions for new interviewers throughout data collection.

Figure 13. The 2004 National Postsecondary Student Aid Study (NPSAS:04) error rate: question delivery



SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Figure 14. The 2004 National Postsecondary Student Aid Study (NPSAS:04) error rate: data entry



SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Quality circle meetings

Quality circle meetings provided an opportunity for NPSAS:04 interviewers to discuss data collection issues with project staff. Topics discussed during these meetings covered all aspects of data collection, including help desk, tracing and locating, and interviewing. Meetings were scheduled weekly during the day and evening shifts to ensure that all telephone interviewers had an opportunity to attend. Summaries of the discussions and decisions addressed during these meetings were compiled and distributed to all interviewers in the form of a newsletter. Issues covered in quality circle meetings included problem sheets, coding strategies, achieving gatekeeper cooperation, interview logic, and clarification of the intent of questions and help text.

This page left intentionally blank.

Chapter 5
Variable Construction and File Development

The data files for the 2004 National Postsecondary Student Aid Study (NPSAS:04) contain student-level and institution-level data collected from institution records, government databases, and student interviews. These files are available as a set of restricted research files, fully documented by an electronic codebook (ECB), and as a public release Data Analysis System (DAS), which also contains full documentation.[44] This chapter describes each file and details the editing and documentation process.

5.1 Overview of the NPSAS:04 Data Files

The primary analysis file, from which the study DASs were constructed, contains data for approximately 90,700 study respondents. The primary analysis file contains over 500 variables, developed from multiple sources (see table 23 for information on the completeness of data available for study respondents). Throughout the data collection period, data were processed and examined for quality control purposes. Editing of student data began shortly after the start of self-administered Web data collection, when procedures and programs for this purpose were first developed. Similarly, editing of the institution record data began shortly after computer-assisted data entry (CADE) data collection was initialized. Anomalous values were investigated and resolved, where appropriate, through the use of data corrections and logical recodes. Interim files were delivered to the National Center for Education Statistics (NCES) for review throughout the data collection period.

Following completion of all study data collection, separate DAS files were created for undergraduate and graduate/first-professional students. The first DASs, both undergraduate and graduate/first-professional, were adjudicated and approved for public release in February 2005.

Complete data for NPSAS:04 are located on the restricted access files and are documented by the ECB. The restricted files and the ECB are available to researchers who have applied for and received authorization from NCES to access restricted research files. Authorization may be obtained by contacting the NCES Data Security Office. The restricted use NPSAS:04 ECB contains information about the following files:

- NPSAS Analysis File—Contains analytic variables derived from all NPSAS data sources as well as selected direct student interview variables.
- Student Base Data File—Contains raw data collected from institutional records and the student interview for the study respondents.
- Student Interview School Data File—Contains institution data obtained from the student interview for all study respondents. It is a student-level file; however, a student can have more than one record in the file. There is a separate record for each

[44] The electronic codebook (ECB) and Data Analysis System (DAS) are both fully documented software products available from the National Center for Education Statistics (NCES). The DAS is available online at http://nces.ed.gov/das.

student for each postsecondary institution the student attended during the study year (up to six institutions).

- Institution File—Contains selected institution-level variables for the sampled institutions. This file can be linked to the Student Base Data File by the Integrated Postsecondary Education Data System (IPEDS) UNITID number.
- CPS 2003–04 Data File—Contains data received from the Central Processing System (CPS) for the study respondents who matched to the 2003–04 financial aid application files.
- CPS 2004–05 Data File—Contains data received from CPS for the study respondents who matched to the 2004–05 financial aid application files.
- NSLDS Federal Pell Data File—Contains raw grant-level data received from the National Student Loan Data System (NSLDS) for the study respondents who received Pell Grants during the NPSAS year or prior years. This is a history file with separate records for each transaction in the Pell system.
- NSLDS Federal Loans Data File—Contains raw loan-level data received from NSLDS for the study respondents who received loans during the NPSAS year or prior years. This is a history file with separate records for each transaction.
- Weights File—Contains all the sampling and analysis weights created for NPSAS:04. There is a separate record for each study respondent.
- Weight History File—Contains all intermediate weight adjustment factors, as well as the final institution and student weights created for NPSAS:04. There is a separate record for each study respondent.

5.2 Online Coding and Verification

5.2.1 Online Coding

The web-based student interview included an online coding system used to obtain IPEDS information for postsecondary institutions (other than the NPSAS institution from which they were sampled) that the student attended during the study year. After providing the state and city in which the institution is located, the online coding system displayed the list of all postsecondary institutions in that location, and the respondent or interviewer could select the appropriate institution. Upon selection, the name of the institution, as well as selected IPEDS variables (institutional level, control) were inserted into the database. This online coding system greatly reduced the IPEDS coding effort and amount of IPEDS file merging necessary after data collection was over.

5.2.2 Range and Consistency Checks

NPSAS:04 included two major web-based data collection systems: student record abstraction and the student interview. Both systems included edit checks to ensure data collected were within valid ranges. To the extent feasible, both systems incorporated across-item consistency edits. Whereas more extensive consistency checks would have been technically possible, use of such edits was limited to prevent excessive respondent burden. Below is a

description of the online range and consistency checks incorporated into the two Web instruments.

General verifications

- Range checks were applied to all numerical entries, such that only valid numeric responses could be entered.
- If, in response to a "check all that apply" question, a valid answer and the "none of the above" option were both checked, respondents and interviewers were notified to uncheck other options before checking the "none of the above" option.
- Pop-up messages confirmed responses that fell outside prespecified ranges for selected numeric values such as income and hours worked per week. Some checks were soft, allowing the respondent to keep the out-of-range response, and some checks were hard, requiring that the respondent update the response to one that fell within the valid range.
- Consistency checks identified conflicting responses (e.g., if the highest degree expected to earn was lower than the current degree) and allowed respondents the opportunity to change answers as appropriate.

5.3 Data Editing

The NPSAS:04 data were edited using procedures developed and implemented for previous NCES-sponsored studies, including NPSAS:2000. Edit checks were performed on the NPSAS:04 student interview data and CADE data, both during and upon completion of data collection, to confirm that the intended skip patterns were implemented in both instruments. At the conclusion of data collection, special codes were added as needed to indicate the reason for missing data. Missing data within individual data elements can occur for a variety of reasons. Table 45 lists each missing value code and its associated meaning in the NPSAS:04 data files.

Table 45. Description of missing data codes: 2004

Missing data code	Description
	Don't know (student interview)
–1	Data not available (computer-assisted data entry [CADE])
–3	Not applicable
–6	Value out of range
–7[1]	Item was not reached (either partial interviews or student interview nonrespondents)
–8	Item was not reached due to an error
–9	Data missing, reason unknown

[1] This code was only applicable for student interview data items.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Skip-pattern relationships in the database were examined by methodically running cross-tabulations between gate items and their associated nested items. In many instances, gate-nest relationships had multiple levels within the CADE or student instrument. That is, items nested within a gate question may themselves have been gate items for additional items. Therefore, validating the gate-nest relationships often required much iteration and many multiway cross-tabulations.

The data cleaning and editing process for the NPSAS:04 CADE and student interview data involved a multistage process that consisted of the following steps:

Step 1. Blank or missing data were replaced with -9 for all variables in the instrument database. A one-way frequency distribution of every variable was reviewed to confirm that no missing or blank values remained. These same one-way frequencies revealed any out-of-range or outlier values, which were investigated and checked for reasonableness against other data values. Example: hourly wages of $0.10, rather than $10.00. Creating SAS formats from expected values and the associated value labels also revealed any categorical outliers.

Descriptive statistics were produced for all continuous variables. All values less than zero were temporarily recoded to missing. Minimum, median, maximum, and mean values were examined to assess reasonableness of responses and anomalous data patterns were investigated and corrected as necessary.

Step 2. Legitimate skips were identified using instrument source code. Gate-nest relationships were defined to replace -9's (missing for unknown reason) with -3's (not applicable) as appropriate. Two-way cross-tabulations between each gate-nest combination were evaluated, and high numbers of nonreplaced -9 codes were investigated to ensure skip-pattern integrity.

Nested values were further quality checked to reveal instances in which the legitimate skip code overwrote valid data which typically occurred if a respondent answered a gate question and the appropriate nested item(s), but then backed up and changed the value of the gate, following an alternate path of nested item(s). Responses to the first nested item(s) remained in the database and, therefore, required editing.

In cases where it could not be determined whether nested items had been legitimately skipped because the response to the gate item was indeterminate (either blank, -9, or don't know, -1), the edit code replaced -9's in nested items with the same value as the gate item. In this way, the value of the gate item was carried through to the nested items.

Step 3. Variable formatting (e.g., formatting dates as YYYYMM) and standardization of time units, for items which collected amount of time in multiple units, were performed during this step. In addition, any new codes assigned by expert coders reviewing IPEDS codes from the student interview (including those institutions that were unable to be coded during the interview) were merged back with the interview data files.

Also at this step, logical recodes were performed when the value of missing items could be determined from answers to previous questions or preloaded values. For instance, if the student did not work while enrolled, then the amount earned should have been coded to $0 rather than -3 or -9. If a student indicated he or she was not disabled, then the "nested" disability items under the gate question were logically recoded to "no."

Step 4. At this step,[45] special codes of -3 and -9 in the student interview file were replaced with -7 (item not administered) based on the section completion indicators. The -7 code allows analysts to easily distinguish items that were either skipped or simply left blank from items not administered (cases where the respondent broke off during the Web interview, or for study respondents who were nonrespondents to the student interview.)

Step 5. One-way frequency distributions for all categorical variables and descriptive statistics for all continuous variables were examined. Out-of-range or outlier values were either replaced with the value of -6 (bad data, out of range) or recoded to a more reasonable value. For example, in CADE, if a user reported a Pell Grant amount for a student of more than $4,050 (the maximum amount allowed) that value was set to $4,050.

Step 6. One-way frequencies on all categorical variables were regenerated and examined. Variables with high counts of -9 values were investigated. Because self-administered Web respondents could skip over most items without providing an answer, -9's did remain a valid value, especially for sensitive items, such as those asking for financial information.

Concurrent with the data cleaning process, detailed documentation was developed to describe question text, response options, logical recodes, and the "applies to" text for each delivered variable.

5.4 Data Perturbation

To protect the confidentiality of NCES data that contain information about specific individuals, NPSAS:04 data were subject to perturbation procedures to minimize disclosure risk. Perturbation procedures, which have been approved by the NCES Disclosure Review Board, preserve the central tendency estimates but may result in slight increases in non-sampling errors.

In a study like NPSAS, there are multiple sources of data for some variables (CPS, CADE, student interview, etc.) and reporting differences can occur in each. Data swapping and other forms of perturbation, implemented to protect respondent confidentiality, can lead to inconsistencies as well.

5.5 Statistical Imputations

All variables with missing data were imputed, following procedures[46] described by Ault et al. (2003). The imputation procedures employed a two-step process. In the first step, the matching criteria and imputation classes that were used to stratify the dataset were identified such that all imputation was processed independently within each class. In the second step, the weighted sequential hot deck process was implemented,[47] whereby missing data were replaced with valid data from donor records that match the recipients with respect to the matching criteria.

[45] This step was not applicable for student record (computer-assisted data entry [CADE]) data.

[46] The methodology described by Ault et al. (2003), was followed with the exception that variances resulting from imputing variable values were not calculated.

[47] The term "hot deck" refers to the fact that the set of potential donors changes for each recipient. In contrast, cold deck imputation defines one static set of donors for all recipients. In all such imputation schemes, the selection of the donor from the entire deck is a random process.

Variables requiring imputation were not imputed simultaneously. However, some variables that were related substantively were grouped together into blocks, and the variables within a block were imputed simultaneously. Basic demographic variables were imputed first using variables with full information to determine the matching criteria. The order in which variables were imputed was also determined to some extent by the substantive nature of the variables. For example, basic demographics (such as age) were imputed first and these were used to process education variables (such as student level and enrollment intensity) which in turn were used to impute the financial aid variables (such as aid receipt and loan amounts).

For variables with less than 5 percent missing data, the variables used for matching criteria were selected based on prior knowledge about the dataset and the known relationships between variables. For example, in almost all cases student's age and enrollment intensity (full-time/part-time status) were used as matching variables in the imputation process.

For variables with more than 5 percent missing data, a statistical process called Chi-Square Automatic Interaction Detection (CHAID) was used to identify the matching criteria that were most closely related to the variable being imputed (Kass 1980). This step produced a number of imputation classes which contained sets of donors that were used to impute recipients belonging to that class. Imputation classes were formed based on a CHAID analysis of likely candidates for variables related to those being imputed. Efficiency was improved by introducing a common set of related variables as input into the CHAID process (see Ault et al. 2003). The resulting imputation classes varied for each variable or blocks of variables input to CHAID. In the case of the analytically less important variables that were imputed later in the process, such as the raw student interview variables, one common set of imputation classes was used. Efforts were made to define groups of imputation variables for which a common set of imputation classes would be optimal.

Next, the imputation classes were input to a SAS macro that implemented the weighted sequential hot deck procedure. Data were sorted within each imputation class to increase the chance of obtaining a close match between donor and recipient. The hot deck process searches for donors sequentially, starting with the recipient and progressing up and down the sorted file to find the set of eligible donors from which a random selection of one was made. The process is weighted since it incorporates the sample weight of each record in the search and selection routine (Cox 1980; Iannacchione 1982).

In some cases, further intervention was needed to ensure accuracy and consistency of imputation as determined by preexisting edit rules. For example, to impute the level of parents' education, when it is known that the parents have some college but not the parents' specific education level, the potential pool of donors was limited to those with at least some college education, to prevent imputing parents' education level as less than college.

Finally, given the number of variables and the complexity of the relationships among them, it was virtually impossible to identify and eliminate all inconsistencies. The objective was to reduce inconsistencies as much as possible, especially for key analytic variables. The objective of the imputation program was to efficiently impute for all missing data such that the process could be completed within a very short timeframe after the end of data collection. The aim was to replace missing data with data that were valid in all cases, with only a few relatively minor and unimportant exceptions.

Imputation diagnostics consisted of three checks: overall imputation checks, imputation checks by class variables, and multivariate consistency checks. The overall imputation checks compared the sum of the weights and unweighted counts for each level of the imputed variable before and after imputation. The imputation checks by class variables evaluated the number of times a given observation was used as a donor, and compared the sum of the weights and unweighted counts for each level of the imputed variable in the defined imputation classes before and after the imputation. Differences of 5 percent or more flagged the imputation class for further review. Finally, multivariate consistency checks ensured that relationships between variables were maintained and that any special instructions for the imputation were implemented properly.

In any of the three aforementioned checks, if there was any evidence of substantial deviation from the weighted sums or any identified inconsistencies, the imputation process was revised and rerun. For a few variables, the inconsistencies were corrected without rerunning the imputation. In these cases, the inconsistencies were corrected after the imputation.

Some results of the imputation process are provided in appendix H which presents the percentage missing for each variable subject to imputation, both for the total sample and for undergraduate students, as well as pre- and post-imputation distributions for eight key variables.

5.6 Composite and Derived Variable Construction

Analytic variables were created by examining the data available for each student from the various data sources, establishing relative priorities of the data sources—on an item-by-item basis—and reconciling discrepancies within and between sources. In some cases, the derived or composite variables were created by simply assigning a value from the available source of information given the highest priority. In other cases, raw interview items were recoded or otherwise summarized to create a derived variable. A listing of the set of analysis variables derived for NPSAS:04 appears in appendix I. Specific details regarding the creation of each variable appear in the variable descriptions contained in the ECB and DAS.

This page left intentionally blank.

Chapter 6
Unit Nonresponse Bias Analyses, Weighting, and Variance Estimation

Statistical analysis weights were computed for study respondents (defined in section 3.2), so that the study respondents represent the target population described in section 2.1. The statistical analysis weights compensated for the unequal probability of selection of institutions and students in the 2004 National Postsecondary Student Aid Study (NPSAS) sample. The weights also adjusted for multiplicity at the institution and student levels, unknown student eligibility, nonresponse, and poststratification. The institution weight was computed and then used as a component of the student weight. Weights were computed for study respondents as the product of the following 13 weight components:

(1) institution sampling weight (WT1);

(2) institution multiplicity adjustment (WT2);

(3) institution poststratification adjustment (WT3);

(4) institution nonresponse adjustment (WT4);

(5) student sampling weight (WT5);

(6) student subsampling weight (WT6);

(7) first student multiplicity adjustment (WT7);

(8) student unknown eligibility adjustment (WT8);

(9) student not located adjustment (WT9);

(10) student refusal adjustment (WT10);

(11) student other nonresponse adjustment (WT11);

(12) second student multiplicity adjustment (WT12); and

(13) student poststratification adjustment (WT13).

Each weight component, described in the following sections, represents either a probability of selection or a weight adjustment. All nonresponse, extreme weight, and poststratification adjustments were computed using RTI's proprietary generalized exponential models (GEM) (Folsom and Singh 2000), which are similar to logistic models using bounds for adjustment factors and bounds on variance inflation. The GEM approach is a general version of weighting adjustments based on Deville and Särndal's logit model (1992). GEM is not a competing method to weighting class adjustment, rather it is a method utilized to do weight adjustments with a choice of optional features to employ. GEM controls at the margins as opposed to controlling at the cell level, as with weighting class adjustments. This allows consideration of greater numbers of variables. GEM is designed so that the sum of the unadjusted weights for all eligible units equals the sum of the adjusted weights for the respondents. GEM also constrains the nonresponse adjustment factors to be greater than or equal to one.

To prevent the variance from becoming too large, the bounds on adjustment factors were loosened, where necessary. The unequal weighting effects (UWEs) and maximum adjustment factors were monitored to ensure reasonable values.

A key feature and advantage of the GEM software is that the nonresponse adjustment and weight trimming and smoothing are all accomplished in one step. Lower and upper bounds are set on the weight adjustment factors. The bounds on the weight adjustment factors can vary, depending on whether the weight falls inside or outside a range, such as the one defined by the bounds used to identify extreme weights (median weight ± 3 times the interquartile range). This allows different bounds to be set for adjustments for weights that are considered high extreme (weight = median +3 times the interquartile range), low extreme (weight = median -3 times the interquartile range), or non-extreme. In this way, the extreme weights can be controlled and the design effect due to unequal weighting can be reduced. See appendix J for details of the GEM procedure.

The bias in an estimated mean based on respondents, $\bar{y}_R$, is the difference between this mean and the target parameter, π, i.e., the mean that would be estimated if a complete census of the target population was conducted and everyone responded. This bias can be expressed as follows:

$$B(\bar{y}_R) = \bar{y}_R - \pi$$

The estimated mean based on nonrespondents, $\bar{y}_{NR}$, can be computed if data for the particular variable are available for most of the nonrespondents. The true target parameter, π, can be estimated for these variables as follows:

$$\hat{\pi} = (1-\eta)\bar{y}_R + \eta\bar{y}_{NR}$$

where η is the weighted unit (or item) nonresponse rate. For the variables that are from the frame, rather than from the sample, π can be estimated without sampling error. The bias can then be estimated as follows:

$$\hat{B}(\bar{y}_R) = \bar{y}_R - \hat{\pi}$$

or equivalently

$$\hat{B}(\bar{y}_R) = \eta(\bar{y}_R - \bar{y}_{NR}).$$

This formula shows that the estimate of the nonresponse bias is the difference between the mean for respondents and nonrespondents multiplied by the weighted nonresponse rate.

Nonresponse bias analysis was conducted when the response rate at any level (institutions, students, items) was below 85 percent.[48] Institution and student nonresponse bias analyses were performed and are described in sections 6.1 and 6.2, respectively. An item nonresponse bias analysis was also performed and is described in section 6.3. Section 6.4 discusses variance estimation, including Taylor series, bootstrap replicate weights, and variance approximation using design effects.

[48] See National Center for Education Statistics (NCES) statistical standards for a discussion of nonresponse bias analysis (U.S. Department of Education 2003).

6.1 Institution Nonresponse Bias Analysis and Weighting

6.1.1 Initial Institution Weight Components

There were two initial institution weight components, described below.

(1) Institution Sampling Weight (WT1)

The sampling weight for each sample institution was the reciprocal of its probability of selection. As described in appendix B, the probability of selection for institution *i* was

$$\pi_r(i) = \begin{cases} \dfrac{n_r S_r(i)}{S_r(+)} & \text{for noncertainty selections} \\ 1 & \text{for certainty selections.} \end{cases}$$

where

n_r = the sample size in stratum r,

$S_r(i)$ = the measure of size for the i-th school in stratum r, and

$Sr(+)$ = the total measure of size for all schools in stratum r.

Therefore, the institution sampling weight was assigned as follows:

$WT1 = 1 / \pi_r (i)$.

(2) Institution Multiplicity Adjustment (WT2)

Each institution on the sampling frame initially had one chance of selection. However, the lists for some sample institutions came from a system office or a main campus. Such lists contained students from more than one institution. Some of these lists clearly identified the campus that each student attended, and each campus was treated as a separate institution.

If a student attended more than one institution or campus, then the student had multiple chances of selection. Student multiplicity adjustments are described below. In NPSAS:04, about 10 enrollment lists were provided that represented more than one institution without clearly identifying which institution or campus each student attended. Therefore, the sample of students was selected from the one list. These institutions were treated as having multiple chances of being selected into the sample because each institution was sampled individually but also was brought into the sample by another institution or campus.

When an institution had two chances of selection, a multiplicity adjustment was performed by first estimating, as if the selections were independent, the probability that either record could be selected:

P(A or B) = P(A) + P(B) - P(A)P(B).

Then, the new sampling weight was calculated as the reciprocal of this probability:

NEW_WT1 = 1 / P(A or B).

When an institution had three chances of selection, a multiplicity adjustment was performed by first estimating the probability that any record could be selected:

P(A or B or C) = (P(A) + P(B) +P(C)) - (P(A)P(B) + P(A)P(C) +P(B)P(C)) + P(A)P(B)P(C).

Then, the new sampling weight was calculated as the reciprocal of this probability:

NEW_WT1 = 1 / P(A or B or C).

When an institution had four or more chances of selection, a multiplicity adjustment was performed by first estimating the probability that any record could be selected:

P(A or B or C or D...) ≈ 1 - (1-P(A)) * (1-P(B)) * (1-P(C)) * (1-P(D)) * ….

Then, the new sampling weight was calculated as the reciprocal of this probability:

NEW_WT1 = 1 / P(A or B or C or D…).

Finally, the multiplicity adjustment factor was derived by dividing the new sampling weight by the old sampling weight,

WT2 = NEW_WT1 / WT1,

for the institutions with positive multiplicity, and setting it to unity (1.00) for all other institutions. Hence, the product of WT1 and WT2 equals NEW_WT1 for the institutions with positive multiplicity and equals WT1 for all other institutions.

6.1.2 Assessing Institution Nonresponse Bias

As shown in chapter 3 (table 8), the institution weighted response rate was below 85 percent for all institutions and for six of the nine types of institutions. Therefore, a nonresponse bias analysis was conducted for all institutions and for the six types of institutions with a weighted response rate below 85 percent. A nonresponse bias analysis was also conducted for eight state-level sectors with a weighted response rate less than 85 percent. The nonresponse bias was estimated for variables known, i.e., nonmissing, for most respondents and nonrespondents. There are extensive data available for all institutions from the Integrated Postsecondary Education Data System (IPEDS), and the following variables were used:[49]

- type of institution;[50]
- Carnegie classification;
- degree of urbanization;
- Office of Business Economics (OBE) region;
- historically Black college or university indicator;
- percent of students receiving federal grant aid;
- percent of students receiving state/local grant aid;
- percent of students receiving institutional grant aid;
- percent of students receiving student loan aid;
- percent of students enrolled: Hispanic;

[49] For the continuous variables, categories were formed based on medians, quartiles, or logical breaks.
[50] Type of institution was only used in the nonresponse bias analysis for all institutions.

- percent of students enrolled: Asian or Pacific Islander;
- percent of students enrolled: Black, non-Hispanic;
- total undergraduate enrollment;
- male undergraduate enrollment;
- female undergraduate enrollment;
- total graduate/first-professional enrollment;
- male graduate/first-professional enrollment; and
- female graduate/first-professional enrollment.

For the institution-level variables listed above, the nonresponse bias was estimated and tested (adjusting for multiple comparisons) to determine if the bias was significant at the 5 percent level. Table 46 shows that about 6 percent of the variable categories are significantly biased for all institutions before nonresponse weight adjustments. When nonresponse bias was evaluated by institution type, the percent of the variable categories with significant bias before nonresponse weight adjustments ranged from 0 to 11 percent. Results of nonresponse bias analysis after weight adjustments are discussed in section 6.1.4.

Table 46. Summary of institution nonresponse bias analysis for all institutions, by type of institution: 2004

Nonresponse bias statistics	All institutions	Public less-than-2-year	Public 2-year	Public 4-year non-doctorate	Private not-for-profit 4-year non-doctorate	Private not-for-profit 4-year doctorate	Private for-profit less-than-2-year
Before weight adjustments							
Mean estimated bias	0.10	0.24	0.14	0.2	0.10	0.19	0.12
Median estimated bias	0.05	0.17	0.08	0.14	0.06	0.06	0.07
Percent significant bias	5.61	6.35	6.85	10.84	2.22	#	4.48
After weight adjustments							
Mean estimated bias	0.13	0.32	0.24	0.25	0.18	0.22	0.22
Median estimated bias	0.05	0.29	0.12	0.23	0.09	0.1	0.19
Percent significant bias	#	#	#	2.41	1.11	#	1.49

Rounds to zero.

NOTE: Nonresponse bias analysis was conducted for all institutions and the six types of institutions with a weighted response rate less than 85 percent.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

6.1.3 Adjusting Institution Weights

There were two additional institution weight components, described below.

(3) Institution Poststratification Adjustment (WT3)

To ensure population coverage, the institution sampling weight adjusted for multiplicity was adjusted to control totals for enrollment by institution type and size using GEM. The enrollment totals came from the 2003 IPEDS fall enrollment file.

Table 47 presents the variables associated with the control totals and the average weight adjustment factors by these variables. The weight adjustment factors from GEM met the following constraints:

- minimum: 0.72;
- median: 1.02; and
- maximum: 1.21.

Table 47. Weight adjustment factors for institution poststratification: 2004

Model predictor variables[1]	Control total[2]	Average weight adjustment factor (WT3)
Total	17,610,549	†
Public less-than-2-year, small	23,644	1.11
Public less-than-2-year, large	54,708	0.96
Public 2-year, small	1,590,649	1.13
Public 2-year, large	4,668,436	0.94
Public 4-year non-doctorate-granting, small	775,807	0.72
Public 4-year non-doctorate-granting, large	1,291,391	1.04
Public 4-year doctorate-granting, small	1,395,624	1.12
Public 4-year doctorate-granting, large	3,235,188	0.98
Private not-for-profit 2-year or less, small	37,930	1.21
Private not-for-profit 2-year or less, large	60,212	0.95
Private not-for-profit 4-year, non-doctorate-granting, small	442,161	1.03
Private not-for-profit 4-year, non-doctorate-granting, large	1,161,689	0.96
Private not-for-profit 4-year doctorate-granting, small	422,843	1.08
Private not-for-profit 4-year doctorate-granting, large	1,406,581	1.02
Private for-profit less-than-2-year, small	101,034	0.95

†Not applicable.
[1] Size for poststratification weighting classes was based on the median enrollment within sector for the institutions on the sampling frame.
[2] Control totals are the sum of enrollment across institutions based on Integrated Postsecondary Education Data System (IPEDS) 2003 enrollment data.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

(4) Institution Nonresponse Adjustment (WT4)

The institutional respondent definition is provided in section 3.1.1. A weighting adjustment using GEM was performed to compensate for nonresponding institutions. The nonresponse adjustments were designed to significantly reduce or eliminate nonresponse bias for variables included in the models. Predictor variables were chosen that were thought to be predictive of response status and were nonmissing for most respondents and nonrespondents. The candidate predictor variables are those used in the nonresponse bias analysis described above with the addition of state.

Predictors used in the nonresponse modeling included all the candidate predictor variables identified as well as certain potentially important interactions. To identify these interactions, the Chi-square automatic interaction detection (CHAID) algorithm (Kass 1980) was used. CHAID is a hierarchical clustering algorithm that successively partitions individuals according to categorical predictors for a categorical dependent variable. The algorithm begins with all study individuals as a whole and cycles over each predictor, finding for each predictor an optimal partition of the individuals according to its levels. The most significant optimal partition is then retained, and the CHAID algorithm is again applied to the members of that partition to

find further partitions using the remaining predictors. The algorithm is stopped after a specified number of partitioning steps or if none of the partitions at a given step is found to be significant.

Application of the CHAID algorithm provided interaction terms for the nonresponse adjustment models. CHAID was run for up to three segments, resulting in identification of two-way and three-way interactions.

Some of the predictor variables (Carnegie classification, female undergraduate enrollment, and graduate/first-professional enrollment) were dropped from the adjustment model due to singularity, which prevents the model from running properly. Singularity occurs when a combination of variables can be used to determine the values of another variable, e.g., total enrollment and male enrollment can be used to determine female enrollment.

Table 48 presents the response rates and the resulting adjustment factors by the model variables. The weight adjustment factors from GEM met the following constraints:

- minimum: 1.00;
- median: 1.08; and
- maximum: 5.10.

Table 48. Weight adjustment factors for institution nonresponse adjustment: 2004

Model predictor variables	Number of respondents	Weighted response rate	Average weight adjustment factor (WT4)
Total	1,360	80.0	†
Institution strata			
Public less-than-2-year	50	74.3	1.32
Public 2-year	320	77.6	1.19
Public 4-year non-doctorate-granting	110	70.3	1.32
Public 4-year doctorate-granting	200	87.1	1.15
Private not-for-profit 2-year or less	70	92.6	1.10
Private not-for-profit 4-year non-doctorate-granting	220	78.1	1.21
Private not-for-profit 4-year doctorate-granting	170	80.8	1.24
Private for-profit less-than-2-year	140	82.3	1.25
Private for-profit 2-year or more	90	88.2	1.14
State			
California	130	65.4	1.46
Connecticut	40	95.8	1.15
Delaware	10	100.0	1.19
Georgia	90	98.6	1.02
Illinois	70	86.8	1.19
Indiana	60	98.3	1.02
Minnesota	70	97.7	1.03
Nebraska	20	81.0	1.37
New York	120	88.7	1.10
Oregon	50	91.9	1.13
Tennessee	70	96.4	1.02
Texas	100	88.9	1.23
Other	540	77.9	1.27

See notes at end of table.

Table 48. Weight adjustment factors for institution nonresponse adjustment: 2004—Continued

Model predictor variables	Number of respondents	Weighted response rate	Average weight adjustment factor (WT4)
Total male graduate/first-professional enrollment[1]			
<= 727	330	75.3	1.24
> 727	310	84.0	1.22
Total Enrollment = 0	730	78.6	1.19
Total graduate/first-professional enrollment[1]			
<= 1,820.5	320	72.6	1.26
> 1,820.5	320	84.9	1.20
Total enrollment = 0	730	78.6	1.19
Total male undergraduate enrollment[2]			
<=281	340	81.8	1.19
>281, <=1251	340	75.7	1.24
>1,251, <=4,208.5	340	77.2	1.21
>4,208.5	350	82.1	1.18
Total undergraduate enrollment[2]			
<= 825.5	330	79.2	1.21
>825.5, <=2,938	340	75.5	1.23
>2,938, <=9,799.5	350	77.8	1.21
>9,799.5	350	82.0	1.18
Percent receiving federal grant aid			
<=25	400	82.3	1.17
>25, <=50	490	77.0	1.19
>50, <=75	230	82.2	1.16
>75	120	79.4	1.36
Unavailable or unknown	120	79.6	1.32
Percent receiving institutional grant aid			
<=25	710	78.7	1.20
>25, <=50	210	83.1	1.17
>50, <=75	130	78.1	1.30
>75	200	82.5	1.16
Unavailable or unknown	120	79.6	1.32
Percent receiving student loan aid			
<=25	430	79.7	1.15
>25, <=50	330	82.9	1.17
>50, <=75	320	82.2	1.17
>75	180	64.0	1.40
Unavailable or unknown	120	79.6	1.32
Percent enrolled: Black, non-Hispanic			
<=25	1,150	80.8	1.22
>25, <=50	130	85.6	1.14
>50, <=75	30	76.8	1.32
>75	50	89.8	1.07
Unavailable or unknown	10	12.7	1.16

See notes at end of table.

Table 48. Weight adjustment factors for institution nonresponse adjustment: 2004—Continued

Model predictor variables	Number of respondents	Weighted response rate	Average weight adjustment factor (WT4)
Percent enrolled: Hispanic			
<=25	1,180	81.6	1.20
>25, <=50	100	76.9	1.22
>50, <=75	30	76.1	1.47
>75	40	88.4	1.24
Unavailable or unknown	10	12.7	1.16
Percent receiving state/local grant aid			
<=25	620	79.2	1.23
>25, <=50	360	76.8	1.21
>50, <=75	180	87.7	1.11
>75	90	86.8	1.10
Unavailable or unknown	120	79.6	1.32
Historically Black college or university			
Yes	30	89.9	1.04
No	1,340	79.8	1.21
Degree of urbanization			
Large city	360	80.6	1.24
Mid-size city	370	81.1	1.18
Urban fringe of large city	260	75.2	1.30
Urban fringe of mid-size city	80	80.7	1.14
Large town	50	85.7	1.15
Small town	170	78.7	1.19
Rural	40	90.0	1.05
Not assigned	30	94.1	1.05
Bureau of Economic Analysis Code (Office of Business Economics [OBE]) Region[3]			
New England	90	67.8	1.32
Mid East	200	74.6	1.32
Great Lakes	210	84.5	1.14
Plains	140	86.6	1.13
Southeast	320	87.3	1.09
Southwest	130	86.0	1.22
Rocky Mountains	40	73.1	1.31
Far West	200	70.3	1.35
Outlying areas	30	99.6	1.00
CHAID segments			
In California	130	65.4	1.46
In one of the 12 states other than California; public, private not-for-profit less-than-4-year or private not-for-profit 4-year non-doctorate-granting; percent receiving institutional grant aid <= 50 percent	380	94.6	1.05
In one of the 12 states other than California; private not-for-profit 4-year doctorate-granting or private for-profit; percent receiving institutional grant aid <= 50 percent	70	82.8	1.22
In one of the 12 states other than California; percent receiving institutional grant aid > 50 and <= 75	70	79.2	1.30

See notes at end of table.

Table 48. Weight adjustment factors for institution nonresponse adjustment: 2004—Continued

Model predictor variables	Number of respondents	Weighted response rate	Average weight adjustment factor (WT4)
In one of the 12 states other than California; percent receiving institutional grant aid > 75	130	90.6	1.09
In one of the 12 states other than California; percent receiving institutional grant aid unavailable or unknown	50	81.4	1.22
Not in one of the 12 states; in New England or Mid East; percent receiving student loan aid <= 75	80	71.4	1.29
Not in one of the 12 states; in New England or Mid East; percent receiving student loan aid > 75 or unavailable or unknown	40	39.9	2.18
Not in one of the 12 states; in region other than New England or Mid East; female graduate/first-professional enrollment > 1,073.5[4]	190	86.6	1.13
Not in one of the 12 states; in region other than New England or Mid East; female graduate/first-professional enrollment = 0[4]	230	77.5	1.21

†Not applicable.
[1] Graduate/first-professional enrollment categories were defined by the median.
[2] Undergraduate enrollment categories were defined by quartiles.
[3] New England = Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont; Mid East = Delaware, District of Columbia, Maryland, New Jersey, New York, Pennsylvania; Great Lakes = Illinois, Indiana, Michigan, Ohio, Wisconsin; Plains = Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota; Southeast = Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, West Virginia; Southwest = Arizona, New Mexico, Oklahoma, Texas; Rocky Mountains = Colorado, Idaho, Montana, Utah, Wyoming; Far West = Alaska, California, Hawaii, Nevada, Oregon, Washington Outlying Areas = American Samoa, Federated States of Micronesia, Guam, Marshall Islands, Northern Mariana, Puerto Rico, Palau, Virgin Islands.
[4] Female enrollment variables were used in Chi-square automatic interaction detection (CHAID) to determine segments but were later excluded from the nonresponse adjustment due to singularities in the model.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

6.1.4 Institution Weighting Adjustment Performance

As shown in table 46, the institution weighting adjustments eliminated some, but not all, significant bias. However, for all institutions, public less-than-2-year institutions, and public 2-year institutions, no significant bias remains after weighting for the variables analyzed. For the other types of institutions, the percent of variable categories with significant bias decreased after weight adjustments. Significant bias was reduced for the variables known for most respondents and nonrespondents, which are considered to be some of the more analytically important variables and are correlated with many of the other variables. Appendix K contains detailed tables showing the estimated bias before and after weight adjustments for each domain for which nonresponse bias analysis was conducted.

Table 49 summarizes the institution weight distributions and the variance inflation due to unequal weighting, i.e., UWE, by institutional type. The median institution weights range from 1.1 for public 4-year non-doctorate-granting institutions to 5.0 for private for-profit less-than-2-year institutions. The mean institution weight ranges from 1.5 for public 4-year doctorate-granting institutions to 11.2 for private for-profit 2-year or more institutions. The UWE is 5.7 overall and ranges from 1.8 for public 4-year doctorate-granting institutions to 6.5 for public less-than-2-year institutions.

Table 49. Institution weight distribution and unequal weighting effects (UWEs): 2004

Analysis domain	Minimum	First quartile	Median	Third quartile	Maximum	Mean	UWE[1]
Total	0.7	1.1	1.5	4.1	152.1	4.6	5.7
Type of institution							
Public less-than-2-year	1.1	1.3	1.8	6.0	123.1	7.7	6.5
Public 2-year	0.9	1.1	1.3	3.1	73.5	3.8	4.5
Public 4-year non-doctorate-granting	0.7	0.9	1.1	2.7	49.9	3.2	4.8
Public 4-year doctorate-granting	1.0	1.0	1.2	1.3	10.7	1.5	1.8
Private not-for-profit 2-year or less	1.0	1.6	2.7	5.6	54.8	5.3	3.1
Private not-for-profit 4-year, non-doctorate-granting	1.0	1.0	1.4	4.6	71.8	4.8	4.2
Private not-for-profit 4-year doctorate-granting	1.0	1.1	1.3	1.8	67.2	2.8	6.4
Private for-profit less-than-2-year	1.1	2.8	5.0	9.1	152.1	8.6	3.8
Private for-profit 2-year or more	1.0	2.4	4.6	10.6	125.8	11.2	4.1

[1] UWE calculated as n $S(Wt)^2$ / $(S\ Wt)^2$.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

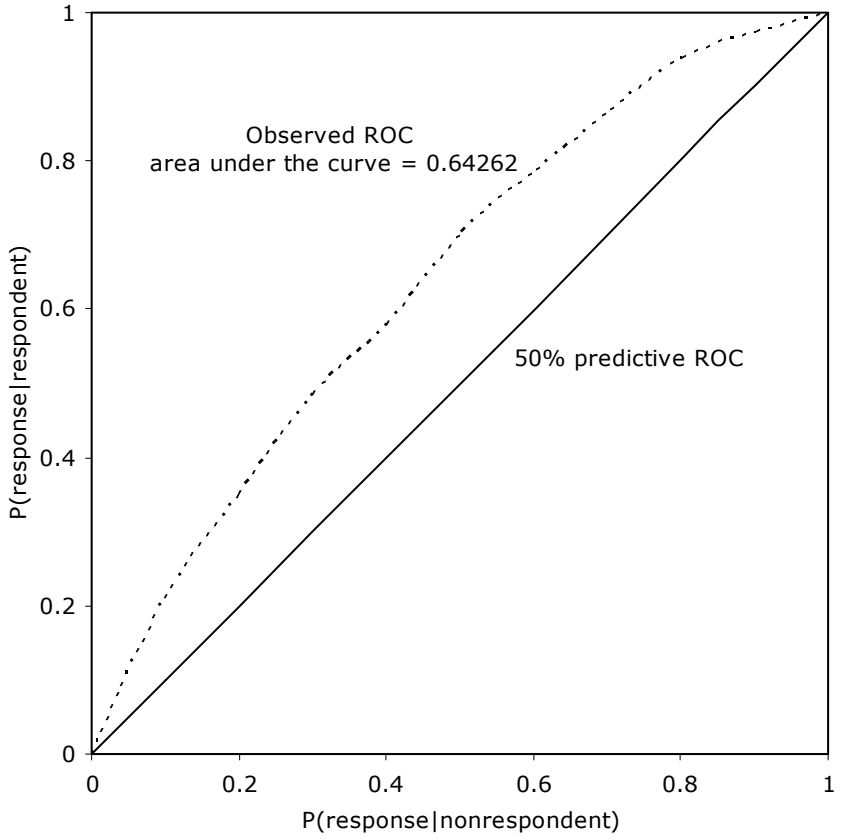
To assess the overall predictive ability of the nonresponse model, a Receiver Operating Characteristics (ROC) curve was used (Hanley and McNeil 1982). The ROC provided a measure of how well the model correctly classified individuals of known response type.[51] The ROC curve was developed in the following manner. For any specified probability, *c*, two proportions were calculated:

- the proportion of respondents with a predicted probability of response greater than *c,* and
- the proportion of nonrespondents with a predicted probability of response greater than *c*.

The plot of the first probability against the second, for *c* from 0 to 1, resulted in the ROC curve shown in figure 15. The area under the curve equals the probability that the fitted model correctly classifies two randomly chosen individuals—one of which is a true respondent and the other a true nonrespondent—where the individual with the higher predicted probability of response is classified as the respondent. An area of 0.5 under an ROC curve indicates that a correct classification is made 50 percent of the time, with the model providing no predictive benefit. An area of 1 indicates that the true respondent always has the higher predicted probability of response, and so the model always classifies the two individuals correctly. Figure 15 shows that the area under the ROC curve is 0.64, so the predicted probabilities give the correct classification 64 percent of the time (about two of every three pairings). Predictive probabilities from ROC curves can also be interpreted in terms of the nonparametric Wilcoxon test statistic, where the ROC area of 0.64 equals the value of the Wilcoxon test statistic. Viewed in this way, the Wilcoxon test provides a significant rejection of the null hypothesis of no predictive ability ($p < 0.05$). This result can be interpreted to mean that the variables used in the model are highly informative but not definitive predictors of a sample institution's overall response propensity.

[51] For a more detailed example of the Receiver Operating Characteristics (ROC) curve use in nonresponse modeling see Iannacchione (2003).

Figure 15. Receiver Operating Characteristics (ROC) curve for overall institution response propensity: 2004



SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

6.2 Student Nonresponse Bias Analysis and Weighting

6.2.1 Initial Student Weight Components

There were four initial student weight components, described below.

(5) Student Sampling Weight (WT5)

The overall student sampling strata were defined by crossing the institution sampling strata with the student strata within institutions. The overall sampling rates for these sampling strata can be found in appendix B. The sample students were systematically selected from the enrollment lists at institution-specific rates that were inversely proportional to the institution's probability of selection. Specifically, the overall stratum sampling rate divided by the institution's probability of selection or

$$f_{s|i} = \frac{f_s}{\pi_r(i)},$$

where f_s = the overall student sampling rate, and

$\pi_r(i)$ = the institution's probability of selection.

As discussed in appendix B, the institution-specific rates were designed to obtain the desired sample sizes and achieve nearly equal weights within the overall student strata.

If the institution's enrollment list was larger than expected based on the IPEDS data, the preloaded student sampling rates would yield larger-than-expected sample sizes. Likewise, if the enrollment list was smaller than expected, the sampling rates would yield smaller-than-expected sample sizes. To maintain control on the sample sizes, the sampling rates were adjusted, when necessary, so that the number of students selected did not exceed by more than 50 students the expected sample size of the institution based on the IPEDS data. A minimum sample size constraint of 10 students also was imposed so that there would be at least four respondents from each participating institution for variance estimation.

The student sampling weight was calculated as the reciprocal of the institution-specific student sampling rates, or

WT5=1/ $f_{s|i}$.

(6) Student Subsampling Weight (WT6)

For paper lists, samples were selected manually, and then the list of sample students was entered into an electronic file. When students from different strata, e.g., first-time beginners (FTBs) and other undergraduates, were combined on a list, the sampling rate from the stratum with the higher rate was used. Then after the sample was entered into an electronic file, the students from the other stratum (or strata) were subsampled.

The student subsampling weight adjustment factor, WT6, was the reciprocal of this subsampling rate. This weight factor was unity (1.00) for most students because this subsampling was not necessary for most institutions.

(7) First Student Multiplicity Adjustment (WT7)

Students who attended more than one eligible institution during the 2003–04 academic year had multiple chances of being selected. That is, they could have been selected from any of the institutions they attended. Therefore, these students had a higher probability of being selected than was represented in their sampling weight.

This multiplicity was adjusted by dividing their sampling weight by the number of institutions attended that were eligible for sample selection. Specifically, the student multiplicity weight adjustment factor was defined as

WT7 = 1 / M,

where M is the multiplicity, or number of institutions attended. The multiplicity was determined from the computer-assisted telephone interview (CATI), the Pell Grant payment file, and the National Student Loan Data System (NSLDS). If student multiplicity was missing, the average number of institutions attended based on students with known number of institutions attended was used. Averages were computed based on type of institution and federal aid receipt.

The weight adjustment factors met the following constraints:

- minimum: 0.03;
- median: 1.02; and

- maximum: 8.32.

(8) Student Unknown Eligibility Adjustment (WT8)

Final eligibility status could not be determined for nonresponding students who were never contacted. These students were treated as eligible, and their weights were adjusted to compensate for the small portion of students who were actually ineligible (as described below).

Weighting classes were defined by the intersection of institution type with the students' matching status to financial aid files (Central Processing System [CPS], Pell, and loan). Table 50 presents the weight adjustment factors applied to the students with unknown eligibility. These weight adjustment factors were based on the estimated rate of eligibility among students with known eligibility status. For the known-eligible students, the weight adjustment factor was set equal to one.

Table 50. Weight adjustment factors for unknown student eligibility status: 2004

Weighting class (institution type, by student type, by matching status to financial aid files)		Number adjusted for unknown eligibility	Weight adjustment factor (WT8)
Total		6,530	†
Public less-than-2-year	Matched Pell or Stafford file	#	1.00
	Matched CPS file only	#	1.00
	No matches	260	0.70
Public 2-year	Matched Pell or Stafford file	180	1.00
	Matched CPS file only	#	1.00
	No matches	3,690	0.81
Public 4-year non-doctorate-granting, undergraduate	Matched Pell or Stafford file	50	1.00
	Matched CPS file only	#	1.00
	No matches	360	0.93
Public 4-year non-doctorate-granting, graduate	Matched Pell or Stafford file	#	1.00
	Matched CPS file only	#	1.00
	No matches	30	0.87
Public 4-year non-doctorate-granting, undergraduate	Matched Pell or Stafford file	30	1.00
	Matched CPS file only	#	1.00
	No matches	570	0.93
Public 4-year non-doctorate-granting, graduate	Matched Pell or Stafford file	10	1.00
	Matched CPS file only	#	1.00
	No matches	210	0.93
Private not-for-profit less-than-4-year	Matched Pell or Stafford file	20	1.00
	Matched CPS file only	#	1.00
	No matches	80	0.66
Private not-for-profit 4-year non-doctorate-granting, undergraduate	Matched Pell or Stafford file	10	1.00
	Matched CPS file only	#	1.00
	No matches	200	0.87
Private not-for-profit 4-year non-doctorate-granting, graduate	Matched Pell or Stafford file	#	1.00
	Matched CPS file only	#	1.00
	No matches	#	0.88
Private not-for-profit 4-year doctorate-granting, undergraduate	Matched Pell or Stafford file	#	1.00
	Matched CPS file only	#	1.00
	No matches	190	0.91
Private not-for-profit 4-year doctorate-granting, graduate	Matched Pell or Stafford file	#	1.00
	Matched CPS file only	#	1.00
	No matches	120	0.91

See notes at end of table.

Table 50. Weight adjustment factors for unknown student eligibility status: 2004—Continued

Weighting class (institution type, by student type, by matching status to financial aid files)		Number adjusted for unknown eligibility	Weight adjustment factor (WT8)
Private for-profit, less-than-2-year	Matched Pell or Stafford file	110	1.00
	Matched CPS file only	#	1.00
	No matches	200	0.46
Private for-profit 2-year	Matched Pell or Stafford file	50	1.00
	Matched CPS file only	#	1.00
	No matches	40	0.63
Private for-profit 4-year, undergraduate	Matched Pell or Stafford file	10	1.00
	Matched CPS file only	#	1.00
	No matches	110	0.83
Private for-profit 4-year, graduate	Matched Pell or Stafford file	#	1.00
	Matched CPS file only/no matches combined	#	0.93

† Not applicable.

Rounds to zero.

NOTE: CPS = Central Processing System.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04)

6.2.2 Assessing Student Nonresponse Bias

As described in section 3.2, a study respondent is defined as any sample member who is determined to be eligible for the study and has valid data from any source for a selected set of key analytical variables. These are minimal data requirements and the vast majority of study respondents were characterized by considerably more complete data.

As shown in table 10, of the 101,010 eligible sample students the unweighted response rate was about 90 percent, and the weighted response rate was 91 percent. The student weighted response rate is also above 85 percent for all types of institutions with the exception of public 2-year institutions. The weighted response rates by type of institution range from about 84 percent for public 2-year institutions to about 97 percent for private not-for-profit 4-year non-doctorate-granting institutions.

Therefore, a nonresponse bias analysis was conducted only for students from public 2-year institutions. A nonresponse bias analysis was also conducted for six state-level sectors with a weighted response rate less than 85 percent. The nonresponse bias was estimated for seven variables known for both respondents and nonrespondents. Five of these variables were known for most sample members, and the remaining two variables were only known for federally aided students. These variables are listed below.

For all sample members:

- region;
- institution total enrollment;
- CPS match (yes/no);
- Pell Grant recipient (yes/no); and
- Stafford Loan recipient (yes/no).

For federally aided students:

- Pell Grant amount; and
- Stafford Loan amount.

Additionally, it was determined that percent part-time fall enrollment and in-state tuition are important variables to include in the nonresponse bias analysis for students in public 2-year institutions. These variables are not known for both respondents and nonrespondents; however, institution-level data available from IPEDS were used to conduct the analyses.

The nonresponse bias was estimated and tested (adjusting for multiple comparisons) for the above variables to determine if the bias was significant at the 5 percent level. Table 51 shows that about 35 percent of the variable categories are significantly biased for students from public 2-year institutions before weight adjustments. Results of the nonresponse bias analysis after weight adjustments will be discussed in section 6.2.4.

Table 51. Summary of student nonresponse bias analysis for all students, in public 2-year institutions: 2004

Nonresponse bias statistics	Public 2-year
Before weight adjustments	
Mean estimated bias	0.11
Median estimated bias	0.04
Percent significant bias	35.42
After weight adjustments	
Mean estimated bias	0.15
Median estimated bias	0.05
Percent significant bias	29.17

NOTE: Nonresponse bias analysis was conducted only for the one type of institution with a weighted response rate less than 85 percent.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

6.2.3 Adjusting Student Weights

There were five additional student weight components, described below. The student weights were further adjusted for nonresponse. The adjustments for nonresponse was performed in three stages because the predictors of response propensity were potentially different at each stage:

- inability to locate the student;
- refusal to be interviewed; and
- other nonresponse.

Using these three stages of nonresponse adjustment achieved greater reduction in nonresponse bias to the extent that different variables were significant predictors of response propensity at each stage.

(9) Student Not Located Adjustment (WT9)

The first type of adjustment for student nonresponse was an adjustment for the inability to locate the student. These weight adjustments were made to compensate for the potential study nonresponse bias.

Predictor variables were chosen that were thought to be predictive of response status and were nonmissing for both study respondents and nonrespondents. The candidate predictor variables included

- institution type;
- in 1 of 12 states with state- representative sample of undergraduates (yes/no);
- region;
- institution enrollment from IPEDS file (categorical);
- student type;
- FTB status;
- Pell Grant receipt (yes/no);
- Pell Grant amount (categorical);
- Stafford Loan receipt (yes/no);
- Stafford Loan amount (categorical);
- Plus Loan amount (categorical);
- federal aid receipt (yes/no);
- CPS record indicator (yes/no);
- Social Security number indicator (yes/no);
- phone number count;
- e-mail address count; and
- mailing address count.

Predictors used in the nonresponse modeling included all the candidate predictor variables identified as well as certain potentially important interactions. CHAID was used to identify these interactions (see description in section 6.1.3). Application of the CHAID algorithm provided interaction terms for each of the nonresponse adjustment models. For each model, CHAID was run for up to three segments, resulting in identification of two-way and three-way interactions. Segments were retained if they were both statistically and practically significant.

The weight adjustments were computed using GEM. The initial model included all of the predictor variables listed above and the interaction segments identified by the CHAID analysis. The model failed to converge with all the variables included, i.e., there was no solution to satisfy all model equations simultaneously. Therefore, a stepwise approach was taken to reduce the variables in the model. In the same step, high-extreme weights were adjusted, truncated, and smoothed by GEM, while the other weights were adjusted for nonresponse.

Table 52 presents the final predictor variables used in GEM to adjust the weights and the average weight adjustment factors resulting from these variables.[52] The weight adjustment factors met the following constraints:

- minimum: 0.20;
- median: 1.00; and
- maximum: 1.00.

Table 52. Weight adjustment factors for student location nonresponse adjustment: 2004

Model predictor variables	Number of located respondents	Weighted response rate	Average weight adjustment factor (WT9)
Total	95,170	95.4	1.07
Type of institution			
Public less-than-2-year	2,340	95.8	1.70
Public 2-year	29,030	91.7	1.10
Public 4-year non-doctorate-granting	8,490	96.7	1.03
Public 4-year doctorate-granting	20,880	97.0	1.03
Private not-for-profit less-than-4-year	2,680	97.4	1.03
Private not-for-profit 4-year non-doctorate-granting	9,120	98.7	1.03
Private not-for-profit 4-year doctorate-granting	9,310	97.7	1.04
Private for-profit less-than-2-year	7,740	96.1	1.07
Private for-profit 2-year or more	5,590	97.8	1.04
Representative state institution			
No	56,880	96.8	1.07
Yes	38,290	93.1	1.08
Bureau of Economic Analysis Code (Office of Business Economics [OBE]) Region[1]			
New England	5,520	96.7	1.05
Mid East	14,630	96.2	1.10
Great Lakes	14,350	96.4	1.05
Plains	7,440	95.1	1.07
Southeast	22,570	96.8	1.05
Southwest	10,410	97.3	1.04
Rocky Mountains	3,760	98.1	1.13
Far West	14,260	89.4	1.13
Outlying Areas, including Alaska and Hawaii	2,230	94.6	1.06
Institution enrollment size[2]			
0 < enrollment total <=1,596	23,550	97.8	1.10
1,596 < enrollment total <=6,567	24,240	96.5	1.05
6,567< enrollment total <=15,397	22,950	94.0	1.09
Enrollment total >15397	24,430	94.9	1.05
Education level			
Undergraduate	83,940	94.9	1.08
Graduate	9,530	97.8	1.03
First-professional	1,700	98.9	1.02
First-time beginner (FTB) status			
FTB	40,370	95.4	1.08
Not FTB	52,210	95.4	1.07
FTB status unknown	2,590	94.9	1.06

See notes at end of table.

[52] See description of the generalized exponential model (GEM) procedure at the beginning of chapter 6.

Table 52. Weight adjustment factors for student location nonresponse adjustment: 2004—Continued

Model predictor variables	Number of located respondents	Weighted response rate	Average weight adjustment factor (WT9)
Pell Grant recipient			
No	65,640	94.2	1.10
Yes	29,530	99.2	1.02
Stafford Loan recipient			
No	60,870	93.5	1.10
Yes	34,300	99.4	1.02
Federal aid recipient			
No	45,940	92.2	1.13
Yes	49,230	99.3	1.02
In Central Processing System (CPS)			
No	38,280	90.4	1.16
Yes	56,890	100.0	1.01
Count of phone numbers			
0	1,670	60.4	1.68
1	46,020	95.3	1.09
2	34,410	97.2	1.04
More than 2	13,080	97.9	1.03
Count of e-mail addresses			
0	31,960	90.0	1.15
1	36,700	96.0	1.05
2	18,460	99.5	1.01
More than 2	8,050	100.0	1.01
Count of mailing addresses			
0 or 1	38,800	92.4	1.12
2	36,000	97.1	1.05
More than 2	20,360	98.7	1.03
Chi-square automatic interaction detection (CHAID) segments			
In CPS	56,890	100.0	1.01
Not in CPS, no preloaded Social Security number (SSN), undergraduate student	7,570	82.0	1.28
Not in CPS, no preloaded SSN, graduate student	1,730	93.5	1.08
Not in CPS, preloaded SSN, undergraduate student	23,400	90.9	1.16
Not in CPS, preloaded SSN, graduate student	5,580	97.4	1.03

[1] New England = Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont; Mid East = Delaware, District of Columbia, Maryland, New Jersey, New York, Pennsylvania; Great Lakes = Illinois, Indiana, Michigan, Ohio, Wisconsin; Plains = Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota;
Southeast = Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, West Virginia; Southwest = Arizona, New Mexico, Oklahoma, Texas; Rocky Mountains = Colorado, Idaho, Montana, Utah, Wyoming; Far West = Alaska, California, Hawaii, Nevada, Oregon, Washington
Outlying Areas = American Samoa, Federated States of Micronesia, Guam, Marshall Islands, Northern Mariana, Puerto Rico, Palau, Virgin Islands.
[2] Enrollment categories were defined by quartiles.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

(10) Student Refusal Adjustment (WT10)

The second stage of the student nonresponse adjustment was an adjustment for refusal, given that the student was located. This additional type of nonresponse adjustment was made to

further compensate for the potential student nonresponse bias. The same GEM procedure was used as in the adjustment for not locating students (WT9). Once again, high-extreme weights were adjusted, truncated, and smoothed by GEM. Candidate predictor variables were the same as those used in the location nonresponse adjustment. As in the location nonresponse adjustment, a CHAID analysis was performed on the predictor variables to detect important interactions.

Table 53 presents the final predictor variables used in GEM to adjust the student weights and the average weight adjustment factor resulting from these variables. The weight adjustment factors met the following constraints:

- minimum: 0.03;
- median: 1.01; and
- maximum: 1.44.

Table 53. Weight adjustment factors for student refusal nonresponse adjustment: 2004

Model predictor variables	Number of nonrefusal respondents	Weighted response rate	Average weight adjustment factor (WT10)
Total	92,690	97.1	1.03
Type of institution			
Public less than 2-year	2,250	94.9	1.12
Public 2-year	27,500	94.7	1.06
Public 4-year non-doctorate-granting	8,310	97.8	1.02
Public 4-year doctorate-granting	20,540	98.3	1.02
Private not-for-profit less-than-4-year	2,650	98.7	1.01
Private not-for-profit 4-year non-doctorate-granting	9,030	98.9	1.01
Private not-for-profit 4-year doctorate-granting	9,170	98.5	1.02
Private for-profit less-than-2-year	7,700	99.2	1.02
Private for-profit 2-year or more	5,540	99.4	1.01
Representative state institution			
No	55,790	97.8	1.03
Yes	36,900	96.1	1.04
Bureau of Economic Analysis Code (Office of Business Economics [OBE]) Region[1]			
New England	5,360	97.4	1.03
Mid East	14,180	97.2	1.04
Great Lakes	14,000	97.5	1.03
Plains	7,180	96.7	1.03
Southeast	22,150	97.9	1.03
Southwest	10,260	98.4	1.02
Rocky Mountains	3,700	98.9	1.01
Far West	13,660	94.2	1.05
Outlying Areas, including Alaska and Hawaii	2,220	99.3	1.01
Institution enrollment size[2]			
0 < enrollment total <=1,596	23,300	98.7	1.03
1,596 < enrollment total <=6,567	23,590	97.5	1.03
6,567< enrollment total <=15,397	22,060	96.5	1.04
Enrollment total >15,397	23,750	96.8	1.03
Education level			
Undergraduate	81,650	97.0	1.03
Graduate	9,360	98.0	1.02
First-professional	1,680	99.3	1.01

See notes at end of table.

Table 53. Weight adjustment factors for student refusal nonresponse adjustment: 2004—Continued

Model predictor variables	Number of nonrefusal respondents	Weighted response rate	Average weight adjustment factor (WT10)
First-time beginner (FTB) status			
FTB	39,500	97.5	1.03
Not FTB	50,700	97.1	1.03
FTB status unknown	2,490	96.5	1.05
Pell Grant recipient			
No	63,230	96.3	1.04
Yes	29,460	99.8	1.01
Stafford Loan recipient			
No	58,460	95.9	1.05
Yes	34,230	99.8	1.01
Federal aid recipient			
No	43,580	95.0	1.06
Yes	49,110	99.8	1.01
In Central Processing System (CPS)			
No	35,800	94.0	1.07
Yes	56,890	100.0	1.00
Count of phone numbers			
0	1,660	99.4	1.00
1	44,410	96.1	1.04
2	33,690	97.7	1.02
More than 2	12,930	98.7	1.02
Count of e-mail addresses			
0	30,550	94.4	1.05
1	35,720	97.1	1.03
2	18,370	99.5	1.01
More than 2	8,050	100.0	1.00
Count of mailing addresses			
0 or 1	36,890	95.0	1.06
2	35,520	98.5	1.02
More than 2	20,280	99.5	1.01
Chi-square automatic interaction detection (CHAID) segments			
In CPS	56,890	100.0	1.00
Not in CPS, no preloaded Social Security number (SSN), no phone number	440	99.1	1.00
Not in CPS, no preloaded SSN, phone number	7,960	90.3	1.12
Not in CPS, preloaded SSN, undergraduate student	21,940	94.2	1.07
Not in CPS, preloaded SSN, graduate student	5,460	97.5	1.02

[1] New England = Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont; Mid East = Delaware, District of Columbia, Maryland, New Jersey, New York, Pennsylvania; Great Lakes = Illinois, Indiana, Michigan, Ohio, Wisconsin; Plains = Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota;
Southeast = Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, West Virginia; Southwest = Arizona, New Mexico, Oklahoma, Texas; Rocky Mountains = Colorado, Idaho, Montana, Utah, Wyoming; Far West = Alaska, California, Hawaii, Nevada, Oregon, Washington
Outlying Areas = American Samoa, Federated States of Micronesia, Guam, Marshall Islands, Northern Mariana, Puerto Rico, Palau, Virgin Islands.
[2] Enrollment categories were defined by quartiles.

NOTE: FTB = First-Time Beginner, SSN = Social Security number, CPS = Central Processing System.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

(11) Student Other Nonresponse Adjustment (WT11)

The third, and final, stage of adjustment for student nonresponse was an adjustment for other study nonresponse, given that the student was located and did not refuse. This additional type of student nonresponse adjustment was made to further compensate for the potential student nonresponse bias. The same GEM procedure was used as in the adjustment for not locating students and student refusals (WT9 and WT10). Candidate predictor variables were the same as those used in the student location and refusal nonresponse adjustments, using a representative state by school-type variable instead of the representative state indicator. The representative state variable was able to be "expanded" for this model without encountering convergence problems, i.e., the model was able to produce adjustment factors with these variables included. As in the other two nonresponse adjustments, a CHAID analysis was performed on the predictor variables to detect important interactions. The resulting segment interactions and all the main effect variables were then included in GEM. High-extreme weights were adjusted, truncated, and smoothed by GEM as in the previous two adjustments.

Table 54 presents the final predictor variables used in GEM to adjust the student weights and the average weight adjustment factor resulting from these variables. The weight adjustment factors met the following constraints:

- minimum: 0.03;
- median: 1.01; and
- maximum: 1.48.

Table 54. Weight adjustment factors for student other nonresponse adjustment: 2004

Model predictor variables	Number of respondents	Weighted response rate	Average weight adjustment factor (WT11)
Total	90,750	97.5	1.02
Type of institution			
Public 4-year or above	2,170	93.1	1.01
Private not-for-profit 4-year or above	26,400	95.3	1.04
Private for-profit 4-year or above	8,160	98.3	1.01
Public 2 year	20,260	98.6	1.01
Private not-for-profit 2-year	2,610	98.0	1.02
Private for-profit 2-year	8,960	99.1	1.01
Public less–than-2-year	9,060	98.9	1.01
Private not-for-profit less than 2-year	7,620	98.2	1.02
Private for-profit less than 2-year	5,500	99.3	1.01
Representative state institution			
All non-representative state institutions	54,950	98.4	1.02
California, public 2-year	2,860	88.3	1.11
California, public 4-year	1,750	97.6	1.02
California, private not-for-profit 4-year	1,230	97.2	1.03
Connecticut, public 2-year	420	94.0	1.05
Connecticut, public 4-year	580	99.9	1.00
Connecticut, private not-for-profit 4-year	540	99.0	1.01
Delaware, public 2-year	460	87.0	1.14
Delaware, public 4-year	570	99.7	1.00
Delaware, private not-for-profit 4-year	480	99.0	1.01

See notes at end of table.

Table 54. Weight adjustment factors for student other nonresponse adjustment: 2004—Continued

Model predictor variables	Number of respondents	Weighted response rate	Average weight adjustment factor (WT11)
Georgia, public 2-year	1,740	98.9	1.02
Georgia, public 4-year	1,100	98.5	1.01
Georgia, private not-for-profit 4-year	520	100.0	1.00
Illinois, public 2-year	1,430	96.8	1.03
Illinois, public 4-year	1,020	98.9	1.01
Illinois, private not-for-profit 4-year	990	98.7	1.01
Indiana, public 2-year	350	99.9	1.00
Indiana, public 4-year	1,080	99.3	1.01
Indiana, private not-for-profit,4-year	660	99.5	1.00
Minnesota, public 2-year	790	87.6	1.12
Minnesota, public 4-year	620	94.3	1.05
Minnesota, private not-for-profit 4-year	580	99.2	1.01
Nebraska, public 2-year	380	99.5	1.03
Nebraska, public 4-year	540	99.4	1.00
Nebraska, private not-for-profit 4-year	340	100.0	1.00
New York, public 2-year	1,700	96.3	1.03
New York, public 4-year	1,550	96.0	1.03
New York, private not-for-profit, 4-year	2,220	99.2	1.01
Oregon, public 2-year	830	98.9	1.01
Oregon, public 4-year	690	98.9	1.01
Oregon, private not-for-profit 4-year	390	99.8	1.00
Tennessee, public 2-year	850	97.0	1.03
Tennessee, public 4-year	800	98.5	1.01
Tennessee, private not-for-profit 4-year	450	98.8	1.01
Texas, public 2-year	2,470	97.3	1.03
Texas, public 4-year	2,060	98.9	1.01
Texas, private not-for-profit 4-year	780	99.7	1.00
Bureau of Economic Analysis Code (Office of Business Economics [OBE]) Region[1]			
New England	5,230	97.8	1.02
Mid East	13,850	97.8	1.02
Great Lakes	13,760	98.2	1.02
Plains	7,000	97.4	1.03
Southeast	21,880	98.5	1.02
Southwest	10,130	98.8	1.01
Rocky Mountains	3,650	98.9	0.99
Far West	13,070	93.7	1.04
Outlying areas, including Alaska and Hawaii	2,180	98.0	1.02
Institution enrollment size[2]			
0 < enrollment total <=1,596	23,050	98.8	1.01
1,596 < enrollment total <=6,567	23,120	98.1	1.02
6,567< enrollment total <=15,397	21,390	96.9	1.03
Enrollment total >15,397	23,180	97.0	1.02
Education level			
Undergraduate	79,840	97.2	1.02
Graduate	9,240	98.9	1.01
First-professional	1,660	99.0	1.01

See notes at end of table.

Table 54. Weight adjustment factors for student other nonresponse adjustment: 2004—Continued

Model predictor variables	Number of respondents	Weighted response rate	Average weight adjustment factor (WT11)
First-time beginner (FTB) status			
FTB	38,800	97.5	1.02
Not FTB	49,520	97.5	1.02
FTB status unknown	2,430	96.7	1.02
Pell Grant recipient			
No	61,360	96.8	1.03
Yes	29,390	99.8	1.00
Pell Grant amount[3] (in dollars)			
Pell Grant amount=0	61,520	96.8	1.03
0 < Pell Grant amount <=1,487	7,310	99.8	1.00
1,487 < Pell Grant amount <=2,500	7,340	99.8	1.00
2,500 < Pell Grant amount <=4,000	7,360	99.9	1.00
Pell Grant amount > 4,000	7,220	99.8	1.00
Stafford Loan recipient			
No	56,600	96.4	1.03
Yes	34,150	99.8	1.00
Federal aid recipient			
No	41,750	95.6	1.04
Yes	48,990	99.8	1.00
Plus amount[4] (in dollars)			
Plus amount=0	86,750	97.4	1.02
0 < Plus amount <=4,764	990	99.3	1.01
4,764 < Plus amount <=7,775	1,010	100.0	1.00
7,775 < Plus amount <=11,700	1,000	100.0	1.00
Plus amount >11,700	1,000	99.5	1.00
In Central Processing System (CPS)			
No	33,860	94.8	1.05
Yes	56,890	100.0	1.00
Count of phone numbers			
0	1,630	97.4	1.01
1	43,170	96.8	1.03
2	33,120	97.9	1.02
More than 2			
Count of e-mail addresses	12,830	99.0	1.01
0	29,430	94.7	1.04
1	34,950	97.7	1.02
2	18,310	99.7	1.00
More than 2	8,050	100.0	1.00
Count of mailing addresses			
0 or 1	35,450	95.7	1.04
2	35,100	98.6	1.01
More than 2	20,200	99.5	1.01
Social Security number (SSN) preloaded			
No	12,860	94.4	1.05
Yes	77,890	98.1	1.02

See notes at end of table.

Table 54. Weight adjustment factors for student other nonresponse adjustment: 2004—Continued

Model predictor variables	Number of respondents	Weighted response rate	Average weight adjustment factor (WT11)
Chi-square automatic interaction detection (CHAID) segments			
In CPS	56,890	100.0	1.00
Not in CPS; undergraduate or first-professional student; public 4-year or private not-for-profit	14,250	91.6	1.09
Not in CPS; undergraduate or first-professional student; Public 2-year or less, private for-profit	13,040	96.5	1.04
Not in CPS; graduate student; private for-profit 2-year or more, public 4-year in Georgia, Indiana, Oregon, or Texas, or private not-for profit 4-year in New York	4,830	98.9	1.01
Not in CPS; graduate student; public 2-year in IL, public 4-year in California, Connecticut, Delaware, or Nebraska, or private not-for-profit 4-year in Connecticut, Delaware, Georgia, Indiana, Minnesota, Nebraska, Oregon, Tennessee, or Texas	770	99.8	1.00
Not in CPS; graduate student; public 4-year in Minnesota or New York, or private not-for-profit 4-year in California	510	93.2	1.07
Not in CPS; graduate student; public 4-year in Illinois or Tennessee, or private not-for-profit 4-year in Illinois	450	96.9	1.03

†Not applicable.

[1] New England = Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont; Mid East = Delaware, District of Columbia, Maryland, New Jersey, New York, Pennsylvania; Great Lakes = Illinois, Indiana, Michigan, Ohio, Wisconsin; Plains; = Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota; Southeast = Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, West Virginia; Southwest = Arizona, New Mexico, Oklahoma, Texas; Rocky Mountains = Colorado, Idaho, Montana, Utah, Wyoming; Far West = Alaska, California, Hawaii, Nevada, Oregon, Washington; Outlying Areas = American Samoa, Federated States of Micronesia, Guam, Marshall Islands, Northern Mariana, Puerto Rico, Palau, Virgin Islands.

[2] Enrollment categories were defined by quartiles.

[3] Pell Grant amounts were defined by quartiles.

[4] Plus amounts were defined by quartiles.

NOTE: FTB = First-Time Beginner, SSN = Social Security number, CPS = Central Processing System.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

(12) Second Student Multiplicity Adjustment (WT12)

An additional adjustment was made to adjust for student multiplicity. This multiplicity adjustment was calculated by dividing the number of institutions attended that were eligible for sample selection (used in the first multiplicity adjustment) by the imputed value for the number of institutions. Specifically, the second student multiplicity weight adjustment factor was defined as

$$WT12 = M / M\_i,$$

where M is the multiplicity, or number of institutions attended, and M_i is the imputed value for multiplicity (see appendix H). M was used in calculating WT7 (the first adjustment for student multiplicity described above), and if the student multiplicity was missing, an average number of students was used. This second adjustment for student multiplicity (WT12) helps correct for underestimating the number of students that only attended one institution.

The weight adjustment factors met the following constraints:

- minimum: 0.23;
- median: 1.00; and
- maximum: 2.00.

(13) Student Postratification Adjustment (WT13)

To ensure population coverage, the student weights were further adjusted to control totals using GEM. Control totals were established for

- amount of Stafford Loans awarded by institution type;
- amount of Stafford Loans awarded by state (for the 12 representative states);
- amount of Pell Grants awarded by institution type;
- amount of Pell Grants awarded by institution type and state (for the 12 representative states);
- non-fall undergraduate enrollment by institution type;
- fall enrollment by institution type; and
- fall enrollment by student type.

The Stafford Loan and Pell Grant control totals were obtained from the Department of Education. The fall enrollment counts were obtained from the 2003 IPEDS Fall Enrollment Survey, and the non-fall enrollment counts were derived from the 2003 IPEDS Fall Enrollment Survey. There were no separate adjustments for extreme weights.

Table 55 presents the variables associated with the control totals and the average weight adjustment factors by these variables. The weight adjustment factors from GEM are summarized below and met the following constraints:

- minimum: 0.51;
- median: 1.16; and
- maximum: 26.83.

After this last weight adjustment was performed, the final student weight (STUDYWT) was computed as the product of the 13 weight components described in this section and in section 6.2.1.

Table 55. Weight adjustment factors for student poststratification: 2004

Model predictor variables	Control total	Average weight adjustment factor (WT13)
Amount of Stafford Loans awarded, by institution type (in dollars)		
Undergraduate students, public less than 2-year	13,026,697,545	1.18
Undergraduate students, public 2-year	7,717,008,637	1.26
Undergraduate students, public 4-year non-doctorate-granting	3,717,049,121	1.76
Undergraduate students, public 4-year doctorate-granting	2,974,409,702	1.30
Undergraduate students, private not-for-profit less-than-4-year	211,125,017	1.28
Undergraduate students, private not-for-profit 4-year non-doctorate-granting	1,888,975,153	1.95
Undergraduate students, private not-for-profit 4-year doctorate-granting	52,213,325	1.14
Undergraduate students, private for-profit less-than-2-year	34,520,490	2.78
Undergraduate students, private for-profit 2-year or more	1,245,598,300	1.87
Graduate/first-professional students, public 4-year non-doctorate-granting	1,172,917,964	2.11
Graduate/first-professional students, public 4-year doctorate-granting	5,580,695,587	1.23
Graduate/first-professional students, private not-for-profit 4-year non-doctorate-granting	2,415,321,110	1.33
Graduate/first-professional students, private not-for-profit 4-year doctorate-granting	6,571,005,610	1.26
Graduate/first-professional students, private for-profit 4-year non-doctorate-granting	1,058,971,758	2.43
Graduate/first-professional students, private for-profit, 4-year doctorate-granting	433,411,394	9.11
Amount of Stafford Loans awarded, by state (in dollars)		
Undergraduate students, California	1,561,080,368	1.23
Undergraduate students, Connecticut	218,611,394	1.06
Undergraduate students, Delaware	65,525,884	1.28
Undergraduate students, Georgia	599,920,776	1.37
Undergraduate students, Illinois	838,754,263	1.22
Undergraduate students, Indiana	612,784,996	1.10
Undergraduate students, Minnesota	582,912,983	1.29
Undergraduate students, Nebraska	197,239,618	1.16
Undergraduate students, New York	1,659,110,944	1.24
Undergraduate students, Oregon	365,006,653	1.23
Undergraduate students, Tennessee	466,198,839	1.24
Undergraduate students, Texas	1,554,631,434	1.23
Amount of Pell Grants awarded, by institution type (in dollars)		
Public less than 2-year	4,307,638,429	1.12
Public 2-year	1,976,176,806	1.16
Public 4-year non-doctorate-granting	683,716,100	1.56
Public 4-year doctorate-granting	4,130,067,523	1.22
Private not-for-profit less-than-4-year	94,928,580	1.07
Private not-for-profit 4-year non-doctorate-granting	780,702,835	1.84
Private not-for-profit 4-year doctorate-granting	40,841,511	0.82
Private for-profit less-than-2-year	39,523,616	1.33
Private for-profit 2-year or more	628,452,073	1.53
Amount of Pell Grants awarded, by institution type for certainty states (in dollars)		
California public 4-year	471,246,772	1.23
California private not-for-profit 4-year	88,579,366	0.91
California public 2-year	574,590,087	1.38
Connecticut public 4-year	19,614,100	1.13
Connecticut private not-for-profit 4-year	15,555,559	0.87
Connecticut public 2-year	20,677,198	1.00
Delaware public 4-year	7,915,051	0.87
Delaware private not-for-profit 4-year	3,709,571	0.81
Delaware public 2-year	6,714,056	4.57
Georgia public 4-year	117,048,105	1.20

See notes at end of table.

Table 55. Weight adjustment factors for student poststratification: 2004—Continued

Model predictor variables	Control total	Average weight adjustment factor (WT13)
Georgia private not-for-profit 4-year	39,698,418	1.11
Georgia public 2-year	109,081,382	1.26
Illinois public 4-year	110,436,966	1.20
Illinois private not-for-profit 4-year	92,229,899	1.05
Illinois public 2-year	168,401,116	1.16
Indiana public 4-year	95,438,161	1.14
Indiana private not-for-profit 4-year	33,005,264	1.12
Indiana public 2-year	65,763,811	1.22
Minnesota public 4-year	53,324,870	1.18
Minnesota private not-for-profit 4-year	25,408,515	1.13
Minnesota public 2-year	76,570,615	1.49
Nebraska public 4-year	26,816,294	1.12
Nebraska private not-for-profit 4-year	13,641,551	1.28
Nebraska public 2-year	22,416,262	1.47
New York public 4-year	293,940,117	1.27
New York private not-for-profit 4-year	254,594,891	1.13
New York public 2-year	247,361,323	1.15
Oregon public 4-year	52,356,782	1.12
Oregon private not-for-profit 4-year	12,010,633	0.84
Oregon public 2-year	64,850,335	1.45
Tennessee public 4-year	84,713,432	1.04
Tennessee private not-for-profit 4-year	38,955,880	1.49
Tennessee public 2-year	79,352,937	1.28
Texas public 4-year	344,812,406	1.14
Texas private not-for-profit 4-year	66,922,625	1.02
Texas public 2-year	422,479,836	1.21
Non-fall enrollment, by institution type		
Public less than 2-year	26,615	0.96
Public 2-year	2,203,978	1.95
Public 4-year non-doctorate-granting	268,489	1.62
Public 4-year doctorate-granting	407,302	1.84
Private not-for-profit less-than-4-year	14,994	1.05
Private not-for-profit 4-year non-doctorate-granting	285,524	2.93
Private not-for-profit 4-year doctorate-granting	67,835	1.20
Private for-profit less-than-2-year	241,908	2.68
Private for-profit 2-year or more	351,043	2.09
Fall enrollment, by institution type		
Public less than 2-year	65,982	0.95
Public 2-year	6,271,184	1.29
Public 4-year non-doctorate-granting	2,156,077	1.13
Public 4-year doctorate-granting	4,572,108	1.06
Private not-for-profit less-than-4-year	94,080	1.25
Private not-for-profit 4-year non-doctorate-granting	1,757,518	1.24
Private not-for-profit 4-year doctorate-granting	1,711,139	0.98
Private for-profit less-than-2-year	266,832	1.31
Private for-profit 2-year or more	764,395	1.22
Fall enrollment, by student type		
Undergraduate	15,186,075	1.18
Graduate	2,134,427	1.01
First-professional	338,813	1.51

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

6.2.4 Student Weighting Adjustment Performance

As shown earlier in table 51, the student weighting adjustments eliminated some, but not all, bias for students in public 2-year institutions. Significant bias was reduced somewhat for the variables known for most respondents and nonrespondents, which are considered to be some of the more analytically important variables and are correlated with many other variables. However, significant bias still remains because there were small numbers of nonrespondents in this type of institution applying for and receiving federal aid. This may be due to the definition of a respondent. All significant bias was eliminated for the non-aid variables, i.e. region, institution total enrollment, percent part-time fall enrollment, and in-state tuition. Appendix K contains detailed tables showing the estimated bias before and after weight adjustments for each domain for which nonresponse bias was conducted.

Table 56 summarizes the institution weight distributions and the variance inflation due to unequal weighting, i.e., UWE, by student type and type of institution. The median student weight ranges from 22 for students in public less-than-2-year institutions to 266 for students in public 4-year non-doctorate-granting institutions. The mean student weight ranges from 42 for students in private not-for-profit less-than-4-year institutions to 322 for students in public 2-year institutions. The UWE is 2.4 overall and ranges from 1.3 for first-professional students to 5.4 for graduate students.

To assess the overall predictive ability of the nonresponse model, an ROC curve was used and developed as described in section 6.1.4. The predicted probabilities of response (c) were obtained as the product of the predicted response probabilities obtained at each of the three GEM nonresponse adjustment steps. Note that for the last two GEM steps (refusal and other nonresponse adjustments), predicted probabilities were not directly available for students who had already been dropped from the model due to nonresponse in an earlier step. For these students, their predicted probability was set equal to the mean of the predicted probabilities of students still in the model.

The plot of the first probability against the second, for c from 0 to 1, resulted in the ROC curve shown in figure 16. The area under the curve equals the probability that the fitted model correctly classifies two randomly chosen individuals—one of which is a true respondent and the other a true nonrespondent—where the individual with the higher predicted probability of response is classified as the respondent. Figure 16 shows that the area under the ROC curve is 0.86, so 86 percent of the time (or close to 9 of every 10 pairings) the predicted probabilities give the correct classification. Predictive probabilities from ROC curves can also be interpreted in terms of the nonparametric Wilcoxon test statistic, where the ROC area of 0.86 equals the value of the Wilcoxon test statistic. Viewed in this way, the Wilcoxon test provides a significant rejection of the null hypothesis of no predictive ability ($p < 0.05$). This level of discrimination implies that the variables used in the model are highly informative but not definitive predictors of a sample student's overall response propensity.

Table 56. Student weight distribution and unequal weighting effects (UWEs): 2004

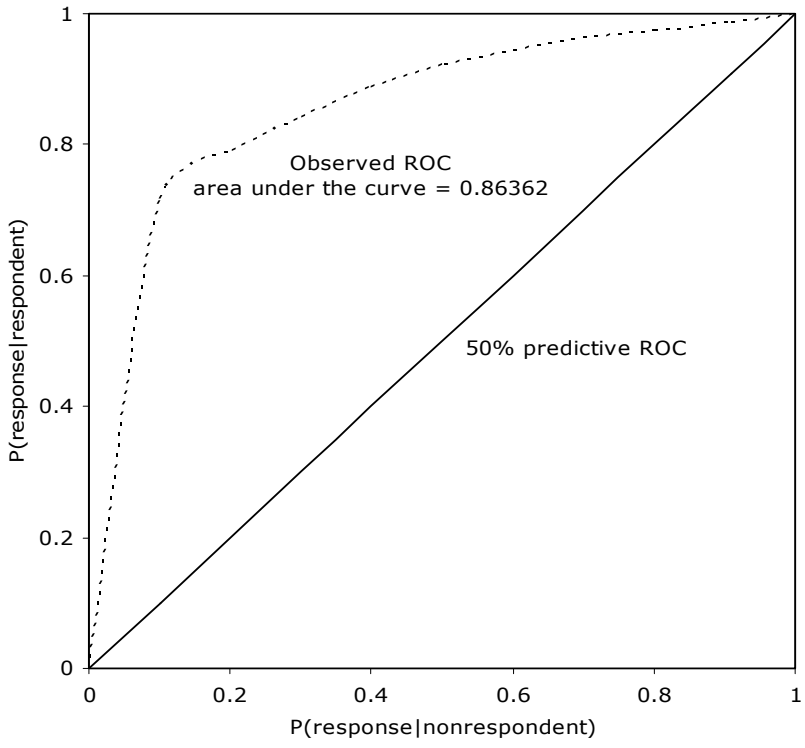
Analysis domain	Minimum	First quartile	Median	Third quartile	Maximum	Mean	UWE[1]
Total	0.2	75.6	166.4	332.6	23468.7	241.1	2.37
Student type							
Undergraduate	0.2	79.2	167.0	331.7	4137.6	238.6	1.97
Graduate	0.6	64.6	118.4	349.1	23468.7	258.0	5.38
First-professional	6.7	193.8	243.3	319.4	2451.9	270.4	1.32
Type of institution							
Public less than 2-year	0.7	11.5	21.7	52.8	745.9	48.1	2.75
Public 2-year	0.9	108.3	201.1	457.8	4137.6	322.0	1.88
Public 4-year non-doctorate-granting	0.8	148.8	266.0	396.3	4220.1	303.5	1.65
Public 4-year doctorate-granting	0.6	116.0	221.9	348.1	2593.3	248.9	1.44
Private not-for-profit less-than-4-year	1.4	16.7	26.1	38.4	1113.2	42.4	2.62
Private not-for-profit 4-year non-doctorate-granting	0.6	85.7	164.1	325.5	23468.7	244.0	5.17
Private not-for-profit 4-year doctorate-granting	0.2	76.8	146.5	274.7	12026.5	195.6	2.10
Private for-profit less-than-2-year	2.0	33.9	52.7	86.7	1362.8	71.2	1.91
Private for-profit 2-year or more	2.4	59.1	106.7	207.4	10216.3	192.0	3.67

[1] UWE calculated as $n \, S(Wt)^2 / (S \, Wt)^2$.

NOTE: UWE = Unequal Weighting Effects.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Figure 16. Receiver Operating Characteristic (ROC) curve for overall student response propensity: 2004



SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

6.3 Item Nonresponse Bias Analysis

When item response rates were less than 85 percent, a nonresponse bias analysis was conducted. Item response rates (RRI) are calculated as the ratio of the number of respondents for whom an in-scope response was obtained (I^x for item x) to the number of respondents who are asked to answer that item. The number asked to answer an item is the number of unit level respondents (I) minus the number of respondents with a valid skip item for item x (V^x). When an abbreviated questionnaire is used to convert refusals, the eliminated questions are treated as item nonresponse (U.S. Department of Education 2003).

$$RRI^x = I^x / (I - V^x)$$

A student is defined to be an item respondent for an analytic variable if that student has data for that variable from any source, including logical imputation. Item response rates were computed using non-imputed data. Valid skips were later logically imputed to the follow-up items after the gate question was imputed. As shown in table 57, the weighted item response rates for all study respondents ranged from about 10 percent to 100 percent. The item response rates by type of institution ranged from about 2 percent to 100 percent.

While values for many variables were derived from multiple sources, including the student interview, student record data, and extant data sources, some variables were obtained from only one source. Given that the weighted response rate to the student interview was about 70 percent, items obtained solely from that source have 30 percent nonresponse even when all interview respondents provided an answer. This issue is compounded for nested items following gate questions, especially those applicable to a small subset of the sample members since follow-up items to unanswered gate items are also treated as nonresponse.

To illustrate an example, the student interview included a set of items about distance education, and was the only source for these data. Students were first asked if they had taken any distance education courses. Those that had were then asked about the types of courses taken. If the first item in the set was not answered, the following questions about the types of distance education courses were treated as nonresponse. More specifically, the gate question (DISTEDUC) had a weighted response rate of about 66 percent, and was therefore missing for about a third of study respondents. Of those who responded to the gate, only about 16 percent reported that they had taken distance education courses. One of the follow-up items, DISTNUM, was not applicable (skipped) for the majority that reported not having taken any distance education courses. These not applicable cases were excluded from the response rate calculation, so the denominator used in computing the response rate for DISTNUM included those cases with a value of ‘yes’ for the gate item (DISTEDUC), as well as those who were nonrespondents to the gate item. Additionally, some students who responded to the gate did not provide a response to the follow-up item, thus DISTNUM has item nonresponse for some cases where DISTEDUC is ‘yes’. Therefore, the low response rate for DISTNUM is driven both by the large amount of missing data for DISTEDUC and the small number of cases where DISTNUM was applicable.

Table 57. Summary of item response rates for all students, by type of institution: 2004

Variable	Variable label	Weighted response rates									
		All students	Public less-than-2-year	Public 2-year	Public 4-year non-doctorate	Public 4-year doctorate	Private not-for-profit less-than-4-year	Private not-for-profit 4-year non-doctorate	Private not-for-profit 4-year doctorate	Private for-profit less-than-2-year	Private for-profit 2-year or more
AGE	Age as of 12/31/03	99.9	100.0	99.9	99.9	100.0	100.0	100.0	100.0	100.0	100.0
AGEGROUP	Age groups as of 12/31/03	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
AIDAPP	Applied for aid	99.4	98.8	98.9	99.7	99.8	99.5	99.9	99.7	99.8	99.9
ATTENDA	Reason for attending NPSAS: complete associate's degree	64.3	58.2	67.1	53.6	47.3	53.9	45.4	35.1	45.6	59.9
ATTENDB	Reason for attending NPSAS: complete certificate	64.3	58.2	67.1	53.6	47.3	53.9	45.4	35.1	45.6	59.9
ATTENDC	Reason for attending NPSAS: learn job skills	64.3	58.2	67.1	53.6	47.3	53.9	45.4	35.1	45.6	59.9
ATTENDD	Reason for attending NPSAS: personal interest or enrichment	64.3	58.2	67.1	53.6	47.3	53.9	45.4	35.1	45.6	59.9
ATTENDE	Reason for attending NPSAS: transfer to 2-year school	64.3	58.2	67.1	53.6	47.3	53.9	45.4	35.1	45.6	59.9
ATTENDF	Reason for attending NPSAS: transfer to 4-year school	64.3	58.2	67.1	53.6	47.3	53.9	45.4	35.1	45.6	59.9
ATTENDG	Reason for attending NPSAS: transfer to another school	64.3	58.2	67.1	53.6	47.3	53.9	45.4	35.1	45.6	59.9
BAYEAR	Year received bachelor's degree	78.4	†	†	84.2	80.4	†	69.0	78.5	†	74.4
CC2000A	Carnegie code (2000) with control	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
CITIZEN2	Citizenship (max non-citizen)	97.5	83.4	95.9	98.9	98.5	99.3	98.4	98.8	97.0	99.2
CLASSA	Type of class: business	55.6	62.3	60.3	55.4	47.3	39.2	48.7	35.1	20.2	24.1
CLASSB	Type of class: health	55.6	62.3	60.3	55.4	47.3	39.2	48.7	35.1	20.2	24.1
CLASSC	Type of class: education	55.6	62.3	60.3	55.4	47.3	39.2	48.7	35.1	20.2	24.1
CLASSD	Type of class: engineering and engineering technology	55.6	62.3	60.3	55.4	47.3	39.2	48.7	35.1	20.2	24.1
CLASSE	Type of class: computer and information sciences	55.6	62.3	60.3	55.4	47.3	39.2	48.7	35.1	20.2	24.1
CLASSF	Type of class: social sciences	55.6	62.3	60.3	55.4	47.3	39.2	48.7	35.1	20.2	24.1
CLASSG	Type of class: natural sciences and mathematics	55.6	62.3	60.3	55.4	47.3	39.2	48.7	35.1	20.2	24.1
CLASSH	Type of class: arts and humanities	55.6	62.3	60.3	55.4	47.3	39.2	48.7	35.1	20.2	24.1
CLASSI	Type of class: communications	55.6	62.3	60.3	55.4	47.3	39.2	48.7	35.1	20.2	24.1
CLASSJ	Type of class: vocational program	55.6	62.3	60.3	55.4	47.3	39.2	48.7	35.1	20.2	24.1
CLASSK	Type of class: university transfer	55.6	62.3	60.3	55.4	47.3	39.2	48.7	35.1	20.2	24.1
CLASSL	Type of class: general education	55.6	62.3	60.3	55.4	47.3	39.2	48.7	35.1	20.2	24.1
CLASSX	Type of class: other	55.6	62.3	60.3	55.4	47.3	39.2	48.7	35.1	20.2	24.1
COMHOUR	Number of hours volunteered per month	38.5	19.5	32.8	40.4	46.4	20.8	43.2	49.5	9.2	28.2
COMONE	One time event	43.3	21.5	36.7	45.3	52.0	23.7	48.3	54.5	10.8	31.8
COMPTO87	Comparable to 1987 NPSAS	99.0	99.7	98.6	99.6	99.4	98.6	99.7	99.8	97.3	97.2
COMREQ	Volunteer work required for graduation/class	42.3	21.6	37.0	46.0	50.8	24.7	50.7	55.6	11.5	28.7
COMSERV	Community service/volunteer in last year	65.5	53.8	64.4	67.2	69.7	49.1	64.5	69.1	41.9	61.6

See notes at end of table.

Table 57. Summary of item response rates for all students, by type of institution: 2004—Continued

Variable	Variable label	Weighted response rates[1]									
		All students	Public less-than-2-year	Public 2-year	Public 4-year non-doctorate	Public 4-year doctorate	Private not-for-profit less-than-4-year	Private not-for-profit 4-year non-doctorate	Private not-for-profit 4-year doctorate	Private for-profit less-than-2-year	Private for-profit 2-year or more
COMSERVA	Volunteer: fundraising (political and non-political)	44.4	22.8	37.9	46.8	53.0	24.9	49.5	55.7	12.1	32.6
COMSERVB	Volunteer: homeless shelter/soup kitchen	44.4	22.8	37.9	46.8	53.0	24.9	49.5	55.7	12.1	32.6
COMSERVC	Volunteer: health services	44.4	22.8	37.9	46.8	53.0	24.9	49.5	55.7	12.1	32.6
COMSERVD	Volunteer: neighborhood improvement	44.4	22.8	37.9	46.8	53.0	24.9	49.5	55.7	12.1	32.6
COMSERVE	Volunteer: service to the church	44.4	22.8	37.9	46.8	53.0	24.9	49.5	55.7	12.1	32.6
COMSERVF	Volunteer: tutoring/education-related	44.4	22.8	37.9	46.8	53.0	24.9	49.5	55.7	12.1	32.6
COMSERVG	Volunteer: other work with kids	44.4	22.8	37.9	46.8	53.0	24.9	49.5	55.7	12.1	32.6
COMSERVX	Volunteer: other	44.4	22.8	37.9	46.8	53.0	24.9	49.5	55.7	12.1	32.6
CONSIDRA	Consider campus safety	63.3	52.3	62.4	65.8	67.5	47.9	64.2	66.5	40.4	60.0
CONSIDRB	Consider graduation rate	63.3	52.3	62.4	65.8	67.5	47.9	64.2	66.5	40.4	60.0
CONSIDRC	Consider job rate	39.4	51.2	†	†	†	24.4	†	†	40.2	†
CRBALDUE	Balance due on all credit cards	29.0	10.2	27.5	32.0	33.5	13.8	27.9	25.1	12.5	22.6
DEGEARN	Earned prior degree/certificates	69.8	61.5	68.9	71.4	73.0	55.9	70.0	73.1	47.1	65.3
DEGEARNA	Already earned bachelor's degree	41.7	32.2	35.8	43.3	47.6	22.1	42.5	55.5	15.7	41.1
DEGEARNB	Already earned associate's degree	41.7	32.2	35.8	43.3	47.6	22.1	42.5	55.5	15.7	41.1
DEGEARNC	Already earned undergraduate certificate/diploma	41.7	32.2	35.8	43.3	47.6	22.1	42.5	55.5	15.7	41.1
DEGEARND	Already earned post-BA certificate	41.7	32.2	35.8	43.3	47.6	22.1	42.5	55.5	15.7	41.1
DEGEARNE	Already earned master's degree	41.7	32.2	35.8	43.3	47.6	22.1	42.5	55.5	15.7	41.1
DEGEARNF	Already earned post-MA certificate	41.7	32.2	35.8	43.3	47.6	22.1	42.5	55.5	15.7	41.1
DEGEARNG	Already earned first professional degree	41.7	32.2	35.8	43.3	47.6	22.1	42.5	55.5	15.7	41.1
DEGEARNH	Already earned doctoral degree	41.7	32.2	35.8	43.3	47.6	22.1	42.5	55.5	15.7	41.1
DEGFIRST	Degree program	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
DELIVE	Distance education: live	22.4	7.2	24.8	23.6	21.0	7.1	22.8	18.0	2.3	25.9
DEPANY	Dependents - has dependents	86.0	72.5	82.2	86.7	87.1	91.4	89.8	88.3	89.1	95.6
DEPCARE	Have dependent children in daycare	64.2	55.4	66.0	65.1	69.9	48.1	61.2	64.7	43.0	60.6
DEPCHILD	Dependents - has dependent children	83.8	69.6	79.7	85.3	86.5	84.6	87.7	87.5	77.3	88.8
DEPCLAIM	Claimed as a dependent	68.4	50.7	67.2	69.9	70.4	51.8	70.2	71.2	43.9	57.9
DEPCOST	Monthly daycare costs	38.2	27.6	38.8	40.6	47.1	22.3	35.1	39.9	17.6	37.3
DEPEND	Dependency status	94.1	93.4	92.6	94.2	93.2	96.2	97.0	96.1	96.7	99.4
DEPINC	Dependent parent income derived	83.9	63.7	77.1	87.1	86.1	88.7	91.7	88.2	81.6	94.7
DEPNUMCH	Dependents - number of dependent children	81.1	65.6	77.3	83.6	85.2	77.1	83.5	86.1	69.6	80.1
DEPOLD	Dependent children - age of oldest	80.4	63.0	76.3	83.1	84.9	74.9	82.5	85.7	68.2	79.2
DEPOTHER	Dependents - has dependent other than children	83.8	69.6	79.7	85.2	86.5	84.6	87.7	87.5	77.3	88.8
DEPYNG	Dependent children - age of youngest	80.4	63.0	76.3	83.1	84.9	74.9	82.5	85.7	68.2	79.2
DERECR	Distance education: pre-recorded	22.4	7.2	24.8	23.6	21.0	7.1	22.8	18.0	2.3	25.9
DEWWW	Distance education: Internet	22.4	7.2	24.8	23.6	21.0	7.1	22.8	18.0	2.3	25.9

See notes at end of table.

Table 57. Summary of item response rates for all students, by type of institution: 2004—Continued

Variable	Variable label	Weighted response rates[1]									
		All students	Public less-than-2-year	Public 2-year	Public 4-year non-doctorate	Public 4-year doctorate	Private not-for-profit less-than-4-year	Private not-for-profit 4-year non-doctorate	Private not-for-profit 4-year doctorate	Private for-profit less-than-2-year	Private for-profit 2-year or more
DISABLE	Disability Flag	65.9	55.5	64.7	67.5	69.7	51.9	65.5	69.7	43.7	62.0
DISMOBIL	Condition that limits physical activities	66.0	55.7	64.8	67.6	69.8	52.0	65.6	69.8	43.8	62.2
DISOTHER	Other condition lasting six months or more	66.0	55.5	64.8	67.7	69.9	52.0	65.6	69.8	43.8	62.1
DISOTHRA	Difficulty: dressing, bathing, etc	10.4	8.2	11.2	10.7	10.9	6.7	9.4	9.3	4.5	9.2
DISOTHRB	Difficulty: getting to school to attend class	10.3	8.1	11.2	10.7	10.9	6.7	9.4	9.2	4.4	9.1
DISOTHRC	Difficulty: learning, remembering	10.4	8.2	11.2	10.7	10.9	6.7	9.4	9.3	4.5	9.2
DISOTHRD	Difficulty: working at a job	10.4	8.2	11.2	10.7	10.9	6.7	9.4	9.3	4.5	9.2
DISSENSR	Have a long-lasting sensory condition	66.1	55.7	64.8	67.7	69.9	52.0	65.6	69.9	43.8	62.3
DISTALL	Distance education: entire program	23.2	7.5	25.5	24.3	22.2	7.5	23.6	18.9	2.5	26.4
DISTEDUC	Distance education: took courses	65.8	54.9	64.8	67.6	70.0	49.5	64.5	69.2	42.3	62.0
DISTLOC	Distance education: location of course(s)	23.1	7.9	25.4	24.3	22.2	7.5	23.4	18.8	2.4	26.4
DISTNUM	Distance education: number of courses	22.3	7.2	24.7	23.4	21.2	7.2	22.4	18.3	2.3	25.7
DISTSATF	Distance education: satisfaction	22.9	7.9	25.1	24.2	22.1	7.4	23.4	18.7	2.4	26.2
DISTYPES	Main limiting condition	16.1	16.3	17.7	16.5	15.3	11.6	15.6	14.2	8.5	16.4
DSTUINC	Dependent student earnings derived	56.6	32.2	42.6	62.5	58.4	77.0	77.3	65.4	72.3	88.4
EMPLWAIV	Tuition waivers for faculty/staff	94.4	96.9	94.1	96.6	94.9	73.4	95.0	96.5	88.0	89.5
EMPLYAM1	Employer tuition aid (excl staff)	98.2	99.2	98.1	99.5	98.5	86.9	98.6	99.2	93.9	95.9
EMPLYAM2	Employer (parents) tuition aid	69.1	59.8	68.4	70.5	72.2	55.2	69.1	71.9	47.7	64.3
ENR01	Monthly enrollment status 2003/07	99.0	99.7	98.6	99.6	99.4	98.6	99.7	99.8	97.3	97.2
ENR02	Monthly enrollment status 2003/08	99.0	99.7	98.6	99.6	99.4	98.6	99.7	99.8	97.3	97.2
ENR03	Monthly enrollment status 2003/09	99.0	99.7	98.6	99.6	99.4	98.6	99.7	99.8	97.3	97.2
ENR04	Monthly enrollment status 2003/10	99.0	99.7	98.6	99.6	99.4	98.6	99.7	99.8	97.3	97.2
ENR05	Monthly enrollment status 2003/11	99.0	99.7	98.6	99.6	99.4	98.6	99.7	99.8	97.3	97.2
ENR06	Monthly enrollment status 2003/12	99.0	99.7	98.6	99.6	99.4	98.6	99.7	99.8	97.3	97.2
ENR07	Monthly enrollment status 2004/01	99.0	99.7	98.6	99.6	99.4	98.6	99.7	99.8	97.3	97.2
ENR08	Monthly enrollment status 2004/02	99.0	99.7	98.6	99.6	99.4	98.6	99.7	99.8	97.3	97.2
ENR09	Monthly enrollment status 2004/03	99.0	99.7	98.6	99.6	99.4	98.6	99.7	99.8	97.3	97.2
ENR10	Monthly enrollment status 2004/04	99.0	99.7	98.6	99.6	99.4	98.6	99.7	99.8	97.3	97.2
ENR11	Monthly enrollment status 2004/05	99.0	99.7	98.6	99.6	99.4	98.6	99.7	99.8	97.3	97.2
ENR12	Monthly enrollment status 2004/06	99.0	99.7	98.6	99.6	99.4	98.6	99.7	99.8	97.3	97.2
EVER2PUB	Ever attended community college	67.6	53.2	100.0	68.2	70.6	49.8	65.5	70.9	43.0	62.9
EVER4YR	Ever attended 4-year school	64.9	58.0	66.6	†	†	53.6	100.0	100.0	45.0	59.8
FEDAPP	Applied for federal aid	96.2	96.8	94.0	95.3	97.8	96.2	98.8	97.3	99.0	99.0
FORESCH	Ever attended elementary or secondary school outside of the U.S.	65.2	53.6	64.5	67.2	68.2	54.0	67.5	69.3	43.8	60.8
FPOFFER	Offered first-professional degree	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

See notes at end of table.

Table 57. Summary of item response rates for all students, by type of institution: 2004—Continued

Variable	Variable label	Weighted response rates[1]									
		All students	Public less-than-2-year	Public 2-year	Public 4-year non-doctorate	Public 4-year doctorate	Private not-for-profit less-than-4-year	Private not-for-profit 4-year non-doctorate	Private not-for-profit 4-year doctorate	Private for-profit less-than-2-year	Private for-profit 2-year or more
GAINSUR	Health insurance with assistantship	77.7	†	†	57.2	80.3	†	47.2	78.9	†	†
GENDER	Gender	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
GPA	Grad point average	89.6	57.9	86.1	95.4	94.7	75.8	92.3	94.2	80.0	86.8
GRADLVL	Graduate class level	95.2	†	100.0	93.3	95.8	†	90.2	96.7	†	100.0
GRADPYR	Year began graduate degree	70.5	†	†	69.7	75.8	†	55.0	70.9	†	67.7
GRADTAA	TA duties: student email	76.2	†	†	59.0	78.1	†	43.3	81.9	†	†
GRADTAB	TA duties: grading	76.2	†	†	59.0	78.1	†	43.3	81.9	†	†
GRADTAC	TA duties: teaching	76.2	†	†	59.0	78.1	†	43.3	81.9	†	†
GRADTAD	TA duties: office hours	76.2	†	†	59.0	78.1	†	43.3	81.9	†	†
GRADTAE	TA duties: discussion	76.2	†	†	59.0	78.1	†	43.3	81.9	†	†
GRADTAF	TA duties: lab	76.2	†	†	59.0	78.1	†	43.3	81.9	†	†
GRASTUIT	Tuition paid by assistantship	73.8	†	†	53.3	76.4	†	46.6	77.7	†	†
GRENRST	Graduate enrollment status (all years)	73.2	†	†	71.5	78.4	†	61.3	72.6	†	70.8
GRFELAMT	Graduate fellowships/grants/traineeships	99.9	100.0	100.0	99.9	99.8	100.0	99.9	99.5	100.0	100.0
GRGRDAMT	Graduate other assistantship amount	99.9	100.0	100.0	99.9	99.8	100.0	99.9	99.5	100.0	100.0
GRINFEL	Institutional graduate fellowships	99.9	100.0	100.0	99.9	99.8	100.0	99.9	99.5	100.0	100.0
GRRESAMT	Research assistantship amount	99.9	100.0	100.0	99.9	99.8	100.0	99.9	99.5	100.0	100.0
GRTEAAMT	Teaching assistantship amount	99.9	100.0	100.0	99.9	99.8	100.0	99.9	99.5	100.0	100.0
GRTRNAMT	Federal traineeships	99.9	100.0	100.0	99.9	99.8	100.0	99.9	99.5	100.0	100.0
HBCU	Historical Black college indicator	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
HIGHLVEX	Highest level of education ever expected	65.7	54.6	64.1	68.2	70.1	49.2	65.5	69.8	41.0	61.6
HISPANIC	Race-ethnicity: Hispanic or Latino origin	92.7	92.6	92.0	92.2	93.7	85.4	93.7	94.6	86.7	92.3
HISPTYPE	Race-ethnicity: Type of Hispanic origin	89.4	84.8	88.4	89.9	92.0	77.0	90.4	91.6	73.7	88.0
HLOFFER	Highest level of offering at NPSAS institution	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
HOMEDIST	Distance from NPSAS school to home	63.0	53.1	62.4	64.6	67.6	45.9	60.7	66.1	38.0	56.3
HOMEPAR	Parents own home	66.1	46.6	63.9	68.1	68.9	50.1	68.2	69.8	41.0	55.7
HOMESTUD	Own home	56.3	50.8	56.0	58.7	60.5	40.7	55.1	59.2	29.8	53.5
HSDEG	High school degree type	90.6	89.3	91.9	89.8	90.5	83.9	89.1	87.9	91.2	86.6
HSGRADYY	High school graduation year	87.4	80.3	89.1	88.1	88.2	81.0	85.6	83.7	80.7	80.3
HSTYPE	Type of high school attended	66.9	54.4	66.2	68.5	70.1	53.2	67.8	69.5	44.6	63.1
INATHAMT	Athletic scholarship	94.9	96.9	94.1	97.0	95.8	73.4	95.9	98.5	88.0	89.5
INDEPINC	Independent student & spouse income derived	81.5	69.0	77.2	81.0	82.8	90.1	84.3	83.9	87.2	94.5
INLNAMT	Institutional loan	94.4	96.9	94.1	96.6	95.1	73.4	95.0	96.5	88.0	89.5
INNSLDS	Positive value in NSLDS 2003-2004 data	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
INPELL	Positive value in Pell data	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
INSMERIT	Institutional merit grants	94.9	96.9	94.1	97.0	95.9	73.4	95.9	98.5	88.0	89.5

See notes at end of table.

Table 57. Summary of item response rates for all students, by type of institution: 2004—Continued

Variable	Variable label	Weighted response rates[1]									
		All students	Public less-than-2-year	Public 2-year	Public 4-year non-doctorate	Public 4-year doctorate	Private not-for-profit less-than-4-year	Private not-for-profit 4-year non-doctorate	Private not-for-profit 4-year doctorate	Private for-profit less-than-2-year	Private for-profit 2-year or more
INSTNEED	Institutional need-based grant	94.9	96.9	94.1	97.0	95.9	73.4	95.9	98.5	88.0	89.5
INSTWRK	Institutional work-study	98.3	99.3	98.1	99.5	98.6	87.0	98.6	99.2	93.8	95.9
INSWAIV	Institutional tuition and fee waivers	94.4	96.9	94.1	96.6	95.1	73.4	95.0	96.5	88.0	89.5
JOBAFFOR	Afford school without working	48.1	29.5	46.5	53.9	55.5	33.6	43.6	48.2	20.9	34.7
JOBCLASS	Job related to coursework	56.9	72.2	61.9	53.2	45.5	31.7	49.0	39.9	46.9	45.3
JOBEARN	Total amount earned during the school year	54.1	42.9	54.1	56.6	56.8	38.5	53.7	53.4	29.6	52.4
JOBEFFA	Job helped with career preparation	48.2	29.5	46.7	54.2	55.6	33.6	43.7	48.3	21.1	34.8
JOBEFFB	Job helped with coursework	48.2	29.5	46.7	54.2	55.6	33.6	43.7	48.3	21.1	34.8
JOBEFFC	Job restricted class choice	48.2	29.5	46.7	54.2	55.6	33.6	43.7	48.3	21.1	34.8
JOBEFFD	Job limited class schedule	48.2	29.5	46.7	54.2	55.6	33.6	43.7	48.3	21.1	34.8
JOBEFFE	Job limited facility access	48.2	29.5	46.7	54.2	55.6	33.6	43.7	48.3	21.1	34.8
JOBEFFF	Job limited number of classes	48.2	29.5	46.7	54.2	55.6	33.6	43.7	48.3	21.1	34.8
JOBEFFGR	Effect of job on grades	66.3	50.5	64.4	69.3	70.3	54.5	65.9	68.4	40.0	59.9
JOBEMPL	Type of employer	58.5	44.9	58.9	60.8	61.3	40.8	57.5	58.1	30.3	56.7
JOBEXPT	Parents expect you to have a job	51.5	31.2	49.4	56.8	58.0	33.7	50.7	52.5	21.2	37.6
JOBHOUR	Hours worked weekly during the school year	59.9	48.4	61.0	61.8	61.5	43.7	59.3	58.1	32.6	58.6
JOBMAIN	Main reason for working	48.2	29.3	46.8	54.2	55.5	33.6	43.7	48.3	21.1	34.6
JOBMAJOR	Job related to major	53.4	26.4	45.7	60.0	61.7	37.2	57.2	58.8	20.6	55.8
JOBNUM	Number of jobs during NPSAS year	68.7	60.7	68.0	70.1	71.9	54.8	68.5	71.5	46.3	64.3
JOBONOFF	Job on or off campus	59.8	46.5	60.6	62.1	62.3	42.7	58.5	58.9	31.7	58.4
JOBPRIOR	Had job prior to enrollment at NPSAS	59.8	46.6	60.5	62.0	62.2	42.6	58.5	58.7	31.6	58.4
JOBROLE	Working student/employee taking classes	61.0	49.5	61.8	63.4	62.8	44.2	61.0	59.6	33.2	59.4
JOBSAVE	Amount saved from summer earnings	55.2	39.4	51.9	59.8	60.7	35.7	58.7	60.1	26.4	45.4
JOBSCHA	Combine school and work: class outside work	36.4	32.2	40.2	32.2	28.7	19.3	39.9	33.3	16.5	46.7
JOBSCHB	Combine school and work: distance ed	36.4	32.2	40.2	32.2	28.7	19.3	39.9	33.3	16.5	46.7
JOBSCHC	Combine school and work: modify schedule	36.4	32.2	40.2	32.2	28.7	19.3	39.9	33.3	16.5	46.7
JOBSUMMR	Work during summer 2003	64.4	51.7	62.6	67.0	68.6	47.9	65.8	68.2	40.5	58.3
JOBWEEK	Weeks worked while enrolled	59.7	46.9	60.5	62.0	62.1	42.8	58.5	58.8	32.2	58.2
LOCALRES	Housing	82.7	66.1	75.0	84.3	86.6	75.0	90.0	92.2	81.5	92.5
MAJORS	Field of study/major (detailed)	90.6	84.5	82.9	95.0	95.4	82.1	96.1	97.3	82.1	94.3
MILTYPE	Military service type	84.7	68.9	80.6	85.6	86.1	90.6	88.2	87.3	86.9	95.4
NPFIRST	NPSAS was first school attended after high school	67.8	58.8	67.2	70.2	71.1	54.1	69.6	70.9	45.4	63.9
NUMCRED	Number of credit cards in own name	66.3	46.6	64.1	68.4	69.0	50.3	68.2	69.9	41.3	55.6
ORPHAN	Orphan or ward of court	93.2	90.7	91.6	93.1	92.5	95.2	96.2	95.2	94.3	98.9
OTHFDGRT	Other federal grants	94.4	96.9	94.1	96.6	95.1	73.4	95.0	96.5	88.0	89.5
OWNINVST	Own investments, business or farm over $10,000	56.5	50.4	56.1	58.3	60.4	40.8	55.2	59.1	32.5	56.1

See notes at end of table.

Table 57. Summary of item response rates for all students, by type of institution: 2004—Continued

Variable	Variable label	Weighted response rates[1]									
		All students	Public less-than-2-year	Public 2-year	Public 4-year non-doctorate	Public 4-year doctorate	Private not-for-profit less-than-4-year	Private not-for-profit 4-year non-doctorate	Private not-for-profit 4-year doctorate	Private for-profit less-than-2-year	Private for-profit 2-year or more
PARALLOW	Monthly allowance amount from parents	75.9	70.4	75.3	76.2	76.6	63.9	75.7	78.5	63.2	79.9
PARBORN	Student's parents were born in the United States	65.3	54.4	64.1	67.0	69.4	49.1	64.2	68.8	41.9	61.6
PARCOLL	Parents taking college courses in 2003-2004	65.8	46.3	63.4	68.0	68.7	49.5	68.1	69.7	40.6	55.6
PARESTA	Parents own other real estate	65.1	45.4	62.9	67.3	67.7	49.5	67.0	68.7	40.3	55.0
PARHELPA	Help from parents: housing	68.4	52.0	67.2	69.9	70.5	51.7	70.1	71.4	44.1	57.9
PARHELPB	Help from parents: other educational expenses	68.4	52.0	67.2	69.9	70.5	51.7	70.1	71.4	44.1	57.9
PARHELPC	Help from parents: other living expenses	68.4	52.0	67.2	69.9	70.5	51.7	70.1	71.4	44.1	57.9
PARHELPD	Help from parents: tuition and fees	68.4	52.0	67.2	69.9	70.5	51.7	70.1	71.4	44.1	57.9
PARLIVE	Lived with parents while not enrolled	69.0	30.4	62.3	70.8	70.9	46.4	74.2	73.2	40.8	56.0
PARPAYCR	Parents help pay credit bills	51.9	22.2	46.7	55.5	57.8	31.5	52.8	57.7	22.6	38.1
PAYOFBAL	Payoff or carry credit balance	51.9	22.2	46.7	55.4	57.8	31.6	52.7	57.8	22.5	38.0
PAYTUIT	Use credit to pay for tuition	52.0	22.4	46.9	55.6	57.8	31.6	53.0	58.0	22.6	38.2
PDADED	Father's highest education level	84.1	70.2	80.3	85.6	86.1	85.3	88.0	86.6	78.8	92.5
PELLAMT	Pell grant amount	99.7	99.8	99.5	99.6	100.0	99.6	99.9	99.9	100.0	99.8
PERKAMT	Perkins loan amount	98.0	98.4	97.0	97.7	98.8	97.6	99.2	98.6	99.7	99.3
PFAMNUM	Dependent student's parent's family size	93.1	91.9	91.1	93.5	92.5	95.6	96.3	95.4	94.8	98.9
PINCOL	Parent's children in college	92.9	91.8	91.0	93.3	92.3	95.6	96.1	95.3	94.7	98.9
PLUSAMT	PLUS loan amount	99.5	99.8	99.2	99.1	100.0	98.5	99.8	99.7	100.0	99.4
PMARITAL	Parent's marital status	93.3	92.1	91.4	93.7	92.6	95.8	96.5	95.5	94.9	99.1
PMOMED	Mothers highest education level	84.2	70.4	80.4	85.4	86.2	85.1	88.2	86.7	78.8	92.3
PRIMLANG	English as primary language	65.7	54.7	64.6	67.4	69.9	49.2	64.4	69.2	41.9	61.6
PRIVLOAN	Private sources (alternative) loans	98.2	99.0	98.1	99.5	98.5	86.7	98.6	99.2	93.6	95.7
PSECTYR	Year first enrolled in postsecondary education	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
RAASIAN	Race--Asian	92.3	94.6	91.9	93.3	94.1	83.0	93.6	93.7	82.2	85.8
RABLACK	Race--Black or African-American	92.3	94.6	91.9	93.3	94.1	83.0	93.6	93.7	82.2	85.8
RAINDIAN	Race--American Indian or Alaska Native	92.3	94.6	91.9	93.3	94.1	83.0	93.6	93.7	82.2	85.8
RAINDTRB	State/federally recognized tribe	63.3	54.1	62.3	62.0	66.4	38.9	62.2	76.8	57.1	61.7
RAISLAND	Race--Native Hawaiian/other Pacific Islander	92.3	94.6	91.9	93.3	94.1	83.0	93.6	93.7	82.2	85.8
RAOTHER	Race--Other	92.3	94.6	91.9	93.3	94.1	83.0	93.6	93.7	82.2	85.8
RAWHITE	Race--White	92.3	94.6	91.9	93.3	94.1	83.0	93.6	93.7	82.2	85.8
REMEDIA	Took remedial course: English	21.4	7.4	24.2	24.4	20.8	12.6	20.1	18.7	2.2	12.7
REMEDIB	Took remedial course: math	21.4	7.4	24.2	24.4	20.8	12.6	20.1	18.7	2.2	12.7
REMEDIC	Took remedial course: reading	21.4	7.4	24.2	24.4	20.8	12.6	20.1	18.7	2.2	12.7
REMEDID	Took remedial course: study skills	21.4	7.4	24.2	24.4	20.8	12.6	20.1	18.7	2.2	12.7
REMEDIE	Took remedial course: writing	21.4	7.4	24.2	24.4	20.8	12.6	20.1	18.7	2.2	12.7
REMETOOK	Took remedial courses this school year	35.3	18.9	39.1	37.7	33.7	21.5	32.7	29.3	8.9	25.8

See notes at end of table.

Table 57. Summary of item response rates for all students, by type of institution: 2004—Continued

Variable	Variable label	Weighted response rates[1]									
		All students	Public less-than-2-year	Public 2-year	Public 4-year non-doctorate	Public 4-year doctorate	Private not-for-profit less-than-4-year	Private not-for-profit 4-year non-doctorate	Private not-for-profit 4-year doctorate	Private for-profit less-than-2-year	Private for-profit 2-year or more
REMEVER	Ever taken remedial courses	66.0	57.1	65.4	68.4	69.3	52.7	67.5	68.8	44.1	62.3
SECTOR1	Institution sector	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
SEOGAMT	FSEOG amount	94.9	96.9	94.1	97.0	95.8	73.4	95.9	98.5	88.0	89.5
SERNEEDA	Adaptive equipment and technology	15.8	16.1	17.3	16.1	15.2	10.8	15.3	13.5	8.0	16.1
SERNEEDB	Alternative exam formats or additional time	15.8	16.1	17.3	16.1	15.2	10.8	15.3	13.5	8.0	16.1
SERNEEDC	Course substitution or waiver	15.8	16.1	17.3	16.1	15.2	10.8	15.3	13.5	8.0	16.1
SERNEEDD	Readers, note takers, or scribes	15.8	16.1	17.3	16.1	15.2	10.8	15.3	13.5	8.0	16.1
SERNEEDE	Registration assistance or priority class registration	15.8	16.1	17.3	16.1	15.2	10.8	15.3	13.5	8.0	16.1
SERNEEDF	Sign language or oral interpreters	15.6	16.1	17.2	15.9	15.0	10.8	15.0	13.3	7.9	15.6
SERNEEDG	Tutors to assist with homework	15.8	16.1	17.3	16.1	15.2	10.8	15.3	13.5	8.0	16.1
SERNEEDX	Needs: other	15.8	16.1	17.3	16.1	15.2	10.8	15.3	13.5	8.0	16.1
SERRECVA	Service: adaptive equipment and technology	16.1	16.0	17.6	16.6	15.2	11.1	15.3	14.0	8.2	16.4
SERRECVB	Service: alternative exam formats/additional time	16.1	16.0	17.6	16.6	15.2	11.1	15.3	14.0	8.2	16.4
SERRECVC	Service: course substitution or waiver	16.1	16.0	17.6	16.6	15.2	11.1	15.3	14.0	8.2	16.4
SERRECVD	Service: readers, note takers, or scribes	16.1	16.0	17.6	16.6	15.2	11.1	15.3	14.0	8.2	16.4
SERRECVE	Service: registration priority or assistance	16.1	16.0	17.6	16.6	15.2	11.1	15.3	14.0	8.2	16.4
SERRECVF	Service: sign language or oral interpreters	16.1	16.0	17.6	16.6	15.2	11.1	15.3	14.0	8.2	16.4
SERRECVG	Service: tutors to assist with homework	16.1	16.0	17.6	16.6	15.2	11.1	15.3	14.0	8.2	16.4
SERRECVX	Service: other	16.1	16.0	17.6	16.6	15.2	11.1	15.3	14.0	8.2	16.4
SIBCOLB4	Siblings in college before respondent	65.9	46.4	63.5	68.1	68.7	49.6	68.2	69.8	40.8	55.4
SINCOL	Number in college (independent students)	84.6	68.4	80.2	85.7	86.3	90.1	87.9	87.6	88.1	95.2
SJCOMSER	Work-study: community service	67.1	75.7	66.6	71.5	65.8	62.3	69.2	62.7	50.2	73.1
SJHOURS	Work study: hours worked per week	78.7	86.1	78.6	77.6	78.4	62.8	81.7	77.2	57.4	80.0
SJMAJOR	Work study: related to major	67.1	55.7	69.6	70.3	64.0	61.4	69.6	61.8	52.7	76.0
SJONOFF	Work study: on/off campus	66.7	73.4	69.0	70.0	63.4	63.5	68.7	61.3	57.8	76.3
SJSCHOOL	Work study: for school	66.6	71.8	68.6	70.3	63.5	62.9	68.8	61.4	58.4	76.3
SJTUTOR	Work study: tutoring	67.2	76.5	66.7	71.7	66.4	62.3	69.2	62.8	49.7	70.6
SMARITAL	Student's marital status	91.7	87.7	89.1	90.9	92.2	99.4	96.2	93.9	96.5	96.7
SPINCOL	Spouse in college	68.0	53.9	67.7	70.9	73.1	54.3	64.0	72.1	43.1	64.6
SPSINC	Spouse of student earnings derived	30.2	25.4	21.5	29.5	27.0	71.5	43.9	34.3	58.7	70.8
SSISSDI	Receive SSI/SSDI	16.2	16.5	17.8	16.6	15.4	11.1	15.4	14.2	8.3	16.3
STAFSUB	Stafford loan subsidized amount	99.5	99.8	99.2	99.0	100.0	98.5	99.8	99.4	100.0	99.4
STAFUNSB	Stafford loan unsubsidized amount	99.5	99.8	99.2	99.0	100.0	98.5	99.8	99.4	100.0	99.4
STATNEED	State need-based grants	94.4	96.9	94.1	96.6	95.1	73.4	95.0	96.5	88.0	89.5
STLNAMT	State loan total	94.4	96.9	94.1	96.6	95.1	73.4	95.0	96.5	88.0	89.5
STMERIT	State merit only grants	94.4	96.9	94.1	96.6	95.1	73.4	95.0	96.5	88.0	89.5
STNOND1	State non-need grants	94.4	96.9	94.1	96.6	95.1	73.4	95.0	96.5	88.0	89.5

See notes at end of table.

Table 57. Summary of item response rates for all students, by type of institution: 2004—Continued

		Weighted response rates[1]									
Variable	Variable label	All students	Public less-than-2-year	Public 2-year	Public 4-year non-doctorate	Public 4-year doctorate	Private not-for-profit less-than-4-year	Private not-for-profit 4-year non-doctorate	Private not-for-profit 4-year doctorate	Private for-profit less-than-2-year	Private for-profit 2-year or more
STUDMULT	Number of institutions attended	83.6	69.5	78.8	85.0	85.8	87.0	87.9	86.7	84.9	94.6
STWKAMT	State work-study total	97.8	99.2	98.0	99.2	97.9	86.7	97.6	97.6	93.5	95.7
TAXHOPE	Claim Federal Hope scholarship	63.0	54.8	62.9	64.5	66.3	48.9	61.3	64.6	42.1	58.4
TAXLEARN	Claim lifetime learning tax credit	63.5	54.7	63.0	65.2	67.3	48.6	61.9	65.8	42.0	58.5
TAXTUIT	Claim tuition tax deduction	64.4	55.3	63.8	65.9	68.0	49.3	63.3	66.9	42.3	60.4
TFEDWRK	Total federal work-study	97.8	99.2	98.0	99.2	97.9	86.7	97.6	97.6	93.5	95.7
TRANSCRD	Transferred credits to NPSAS	30.5	7.7	21.5	44.3	45.9	29.5	44.9	39.8	18.6	32.0
TRANSFR	Transferred from NPSAS	40.6	31.7	43.7	42.2	34.0	47.4	34.4	34.4	36.5	39.7
TRANSPLN	Plan to transfer from NPSAS	51.3	28.7	48.7	55.6	57.7	30.9	55.8	57.9	19.1	46.6
TRANSTO	Transferred to NPSAS	57.2	34.1	54.0	62.5	63.7	61.8	67.8	57.6	45.8	54.0
UGDEGAA	Associate's degree types	96.7	100.0	96.6	89.4	93.7	97.5	100.0	98.0	99.4	99.5
UNTAXBF	Received untaxed benefits in 2003	65.4	57.3	64.7	66.5	70.8	51.6	62.5	69.5	44.0	63.7
UNTAXBFA	Received child support	30.1	25.2	32.0	30.6	31.0	22.5	28.7	25.3	22.1	28.9
UNTAXBFB	Received disability payments	20.1	21.6	23.8	19.5	15.3	16.1	17.9	11.8	16.1	21.6
UNTAXBFC	Received food stamps	20.1	21.6	23.8	19.5	15.3	16.1	17.9	11.8	16.1	21.6
UNTAXBFD	Received social security benefits	20.0	21.4	23.7	19.3	15.2	16.1	17.7	11.6	16.0	21.3
UNTAXBFE	Received TANF	30.1	25.2	32.0	30.6	31.0	22.5	28.7	25.3	22.1	28.9
UNTAXBFF	Received worker's compensation	20.1	21.6	23.8	19.5	15.3	16.1	17.9	11.8	16.1	21.6
USBORN	Respondent born in the U.S.	65.7	54.8	64.4	67.5	69.5	51.8	64.7	69.4	43.3	62.5
VADODAMT	Federal veteran's and military aid	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0
VETBEN	Federal veteran's benefits	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0
VETERAN	Veteran status	95.4	86.6	94.8	95.3	95.9	97.7	94.6	95.4	96.8	99.2
VOCAPPLY	Ever applied for Voc Rehab services	13.9	11.4	15.0	14.0	13.7	9.0	13.4	13.1	7.1	14.1
VOCHELP	State voc rehab and job training (WIA)	94.4	96.9	94.1	96.6	95.1	73.4	95.0	96.5	88.0	89.5
VOCRECV	Ever received Voc Rehab services	16.2	16.5	17.8	16.7	15.4	11.1	15.4	14.3	8.3	16.4
VOTEEVER	Ever vote	65.5	55.1	64.2	67.3	69.3	51.7	64.5	69.2	43.1	62.5
VOTEREG	Registered to vote	65.5	55.1	64.2	67.3	69.2	51.6	64.6	69.1	43.1	62.5
YEARGRAD	Year began graduate school	73.0	†	†	71.4	78.1	†	60.6	72.5	†	70.8

† Not applicable.

NOTE: Nonresponse bias analysis was conducted only for each item with a weighted response rate less than 85 percent. Nonresponse bias analysis was based on the student-level variables known for both respondents and nonrespondents (described in the assessing student nonresponse bias section above). Note that while values for many variables are derived from multiple sources, including the student interview, student record data, and extant data sources, some variables are obtained from only one source. Given that the weighted response rate to the student interview was about 70 percent, items obtained solely from the student interview have 30 percent nonresponse even when all student interview respondents provided an answer. This issue is compounded for nested items following gate questions. Response rates for items that follow a gate item include nonresponse resulting both from nonresponse to the item in question, and also to missing data for previously unanswered gate items. Consequently, item response rates to the follow-up items are deflated because the item is not applicable for an unknown proportion of the nonrespondents to the gate item.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 National Postsecondary Student Aid Study (NPSAS:04).

Therefore, a nonresponse bias analysis was conducted for all items with a weighted response rate less than 85 percent for all students or for students in a particular sector. The possibility of estimating the degree of bias depends on having some variables that reflect key characteristics of respondents and for which there is little or no missing data. The variables listed above in the student-level bias analysis section were used to compare the item respondents and nonrespondents. Additionally, gender and age group were used because they were known for all study respondents. Also, institution strata were used in analyses of items for all students. These variables are important to the study and are related to many of the items being analyzed for low item response rates. For these items, the nonresponse bias prior to imputation was estimated as described in the beginning of chapter 6 for each of these variables known for most respondents and nonrespondents and tested (adjusting for multiple comparisons) to determine if the bias was significant at the 5 percent level.

Appendix K contains a table (table K-23) using one variable (DEPCHILD) to illustrate the estimated bias before imputation for all students. Similar computations were done for about 200 additional variables with item response rates less than 85 percent for all students or for students in at least one sector. Table K-24 in appendix K summarizes these computations. This table also shows a large range for the percent of variable categories with significant bias across all items analyzed prior to imputation. A byproduct of the imputation (described in section 5.5) is the reduction or elimination of item-level nonresponse bias. Imputation reduces or eliminates nonresponse bias by replacing missing data with statistically plausible values. Missing data and the associated nonresponse bias for variables are usually not ignorable (i.e., the respondents' distribution patterns differ from those in the full population). Therefore, replacing missing data with reasonable values produces imputed sample distributions that resemble full population distributions, thus reducing if not eliminating nonresponse bias. The use of carefully constructed imputation classes, donor-imputee matching criteria, and random hot-deck searches within imputation cells are all designed to ensure that imputed data are in fact plausible and that the nonresponse bias is ignorable within the imputation classes.

To evaluate how well the imputation worked in reducing bias for items with a weighted response rate less than 85 percent for all students, the bias was estimated after imputation. For continuous variables, the estimated bias equals the mean before imputation minus the mean after imputation. For categorical variables, the estimated bias was computed for each category as the percentage of students in that category before imputation minus the percentage of students in that category after imputation. The estimated bias was then tested (adjusting for multiple comparisons) to determine if the bias was significant at the 5 percent level. A categorical variable was deemed to be significantly biased if any of the categories was significantly biased. As shown in tables K-25 and K-26 in appendix K, about 25 percent of the variables analyzed still had significant bias after imputation. The relative bias is greater than 10 percent for about 22 percent of the items with remaining significant bias. Analysts should use caution when using the significantly biased items.

6.4 Variance Estimation

For probability-based sample surveys, most estimates are nonlinear statistics. For example, a mean or proportion, which is expressed as $\Sigma wy/\Sigma w$, is nonlinear because the denominator is a survey estimate of the (unknown) population total. In this situation, the variances of the estimates cannot be expressed in closed form. Two procedures for estimating

variances of survey statistics are the Taylor series linearization procedure and the bootstrap replication procedure, which are both available on the NPSAS data files. The analysis strata and replicates created for the Taylor series procedure are discussed in section 6.4.1, and section 6.4.2 discusses the replicate weights created for the bootstrap procedure. Section 6.4.3 discusses the computation and use of design effects to measure the effects that complex sample design features had on the variances of survey estimates.

6.4.1 Taylor Series

The Taylor series variance estimation procedure is a well-known technique used to estimate the variances of nonlinear statistics. The procedure takes the first-order Taylor series approximation of the nonlinear statistic and then substitutes the linear representation into the appropriate variance formula based on the sample design. Woodruff (1971) presented the mathematical formulation of this procedure.

For stratified multistage surveys, the Taylor series procedure requires analysis strata and analysis primary sampling units (PSUs), also called replicates, defined from the sampling strata and PSUs used in the first stage of sampling. For NPSAS:04, analysis strata and analysis PSUs were defined separately for all students combined and can be used for analyses of any domain.

The first step was to identify the PSUs used at the first stage of sample selection. As discussed in chapter 2, the PSUs included the 860 noncertainty institutions. For the 810 certainty institutions, however, the students represent the first stage of sampling. To obtain appropriate degrees of freedom for variance estimation, the students selected from each certainty institution were partitioned into two, three, or four pseudo-PSUs by random assignment of sample students into approximately equal-sized groups. The number of pseudo-PSUs formed was based on the institution's measure of size for first-stage sampling.

The next step was to sort the PSUs and pseudo-PSUs by the 58 institution strata, then by certainty versus noncertainty, and then by the selection order for the noncertainty institutions and by IPEDS ID for the certainty institutions. From this sorted list, the analysis PSUs were then defined by collapsing the PSUs and pseudo-PSUs as required so each analysis PSU contained at least four respondents. This sample size requirement ensured stable variance estimates. Analysis PSUs were then paired to form analysis strata. Certainty institutions that included three or four pseudo-PSUs were made a single analysis stratum. This process resulted in 1,005 analysis strata. The names of the analysis strata and analysis PSU variables are ANALSTR and ANALPSU, respectively.

The procedure described above may overestimate the variance because it does not always account for the finite population correction (FPC) at the institution stage of sampling. Alternatively, the Taylor series procedure can account for the FPC if the secondary sampling units (SSUs) and PSU counts are considered in addition to the analysis strata and analysis PSUs. These variable names are FANALSTR, FANALPSU, FANALSSU, and PSUCOUNT for the analysis strata, PSUs, and SSUs and the PSU counts, respectively. FANALSTR and FANALPSU differ from ANALSTR and ANALPSU in that for certainty institutions FANALSTR equals the institutional sampling stratum and FANALPSU equals ANALSTR. Also, FANALSSU equals ANALPSU for certainty institutions. For noncertainty institutions, FANALSTR equals ANALSTR and FANALPSU equals ANALPSU. Also, FANALSSU was created by randomly

dividing ANALPSU into two parts for noncertainty institutions. There are 658 analysis strata when taking the FPC into account.

6.4.2 Bootstrap Replicate Weights

The variance estimation strategy was chosen for NPSAS:04 to satisfy the following requirements:

1. recognition of variance reduction due to stratification at all stages of sampling;
2. recognition of effects of unequal weighting;
3. recognition of possible increased variance due to sample clustering;
4. recognition of effects of weight adjustments for nonresponse and for poststratification of selected total estimates to known external totals;
5. satisfactory properties for estimating variances of nonlinear statistics and quantiles as well as for linear statistics;
6. ability to apply finite population corrections at the institution stage of sampling and reflect the reduction in variance due to the high sampling rates in some first-stage sampling strata; and
7. ability to test hypotheses about students based on normal distribution theory by ignoring the finite population corrections at the student level of sampling.

Commonly applied bootstrap variance estimation techniques satisfy requirements 1 through 5. To meet requirements 6 and 7 as well, a methodology and computer software developed by Kaufman (2004) were applied. This methodology allows for finite population correction factors at two stages of sampling. The application of the method incorporated the finite population correction factor at the first stage only where sampling fractions were generally high. At the second stage, where the sampling fractions were generally low, the finite population correction factor was set to 1.00.

The Kaufman methodology was used to develop a vector of bootstrap sample weights which was added to the analysis file. These weights are zero for units not selected in a particular bootstrap sample; weights for other units are inflated for the bootstrap subsampling. The initial analytic weights for the complete sample are also included for the purposes of computing the desired estimates. The vector of replicate weights allows for computing additional estimates for the sole purpose of estimating a variance. Assuming B sets of replicate weights, the variance of any estimate, $\hat{\theta}$, can be estimated by replicating the estimation procedure for each replicate and computing a simple variance of the replicate estimates, as follows:

$$\text{var}(\hat{\theta}) = \frac{\sum_{b=1}^{B} (\hat{\theta}_b^* - \hat{\theta})^2}{B},$$

where $\hat{\theta}_b^*$ is the estimate based on the b-th replicate weight (where b=1 to the number of replicates) and B is the total number of sets of replicate weights. See appendix L for more details of this variance estimation procedure. Once the replicate weights are provided, this estimate can be produced by most survey software packages (e.g., SUDAAN [RTI International 2004]

computes this estimate by invoking the DESIGN=BRR option). See appendix M for an example of SUDAAN code.

The number of replicate weights was set at 64 based on an empirical investigation of the behavior of variance estimates as the number of replicates increased. This investigation showed that the stability of variance estimates improved with increasing numbers of replicates and became fairly stable for most estimates when between 50 and 55 replicate weights were used. For the 64 replicate weights included on the analysis file (BOOTWT01 – BOOTWT64), the poststratification process was repeated so that replicate weight variation did not include components that would be controlled by replication of the entire process in conjunction with the same poststratification process. For several of the replicates, one or two of the control totals could not be met due to model convergence problems (i.e., there was no solution to satisfy all model equations simultaneously.

6.4.3 Variance Approximation

The survey design effect for a statistic is defined as the ratio of the design-based variance estimate over the variance estimate that would have been obtained from a simple random sample of the same size (if that were practical). It is often used to measure the effects that sample design features have on the precision of survey estimates. For example, stratification tends to decrease the variance, but multistage sampling and unequal sampling rates usually increase the variance. Also, weight adjustments for nonresponse (performed to reduce nonresponse bias) and poststratification increase the variance by increasing the weight variation. Because of these effects, most complex multistage sampling designs, like NPSAS:04, result in design effects greater than one. That is, the design-based variance is larger than the simple random sample variance.

Specifically, the survey design effect for a given estimate, $\hat{\theta}$, is defined as

$$Deff(\hat{\theta}) = \frac{Var_{design}(\hat{\theta})}{Var_{srs}(\hat{\theta})}.$$

Also, the square root of the design effect is another measure, which can also be expressed as the ratio of the standard errors, or

$$Deft(\hat{\theta}) = \frac{SE_{design}(\hat{\theta})}{SE_{srs}(\hat{\theta})}.$$

In appendix N, design effect estimates are presented for important survey domains and estimates among undergraduate students, graduate students, and first-professional students to summarize the effects of stratification, multistage sampling, unequal probabilities of selection, and the weight adjustments. These design effects were estimated using SUDAAN and the bootstrap variance estimation procedure described in section 6.4.2 and appendix L. If one must perform a quick analysis of NPSAS:04 data without using one of the software packages for analysis of complex survey data, the design effect tables in appendix N can be used to make approximate adjustments to the standard errors of survey statistics computed using the standard software packages that assume simple random sampling designs. However, one cannot be confident regarding the actual design-based standard errors without performing the analysis

using one of the software packages specifically designed for analysis of data from complex sample surveys. See appendix M for more details concerning the use of such software packages.

Large design effects imply large standard errors and relatively poor precision. Small design effects imply small standard errors and good precision. In general terms, a design effect under 2.0 is low, 2.0 to 3.0 is moderate, and above 3.0 is high. Moderate and high design effects often occur in complex surveys such as NPSAS, and the design effects in appendix N are consistent with those in past NPSAS studies. Unequal weighting causes large design effects and is often due to nonresponse and poststratification adjustments. However, in NPSAS, the unequal weighting is also due to the sample design and different sampling rates between institution strata and also different sampling rates between student strata.

References

Ault, K., Black, S., Chromy, J., Fahimi, M., Siegel, P., Trofimovich, L., Whitmore, R., and Berkner, L. (2003). *Imputation Methodology for the National Postsecondary Student Aid Study: 2004* (NCES 2003–20). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Chromy, J.R. (1979). Sequential Sample Selection Methods. *Proceedings of the American Statistical Association Section on Survey Research Methods of the American Statistical Association*, 401–406.

Chromy, J.R. (1981). Variance Estimators for a Sequential Sample Selection Procedure. In. D. Krewski, R. Platek, and J.N.K. Rao (Eds.), *Current Top IMS in Survey Sampling* (pp. 329-347). New York: Academic Press.

Cohen, S.B. (1997). An Evaluation of Alternative PC-Based Software Packages Developed for the Analysis of Complex Survey Data. *The American Statistician, 57*(13): 285-292.

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

Deville, J.C., and Särndal, C.E. (1992). Calibration Estimating in Survey Sampling. *Journal of the American Statistical Association, 87*:376-382.

Education Sciences Reform Act of 2002. (2002). Public Law 107-279, Title I, Section 153. Available: http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=107_cong_public_laws&docid=f:publ279.107.pdf.

Hanley, J.A., and McNeil, B.J. (1982). The Meaning and Use of the Area under a Receiver Operating Characteristic Curve. *Diagnostic Radiology, 143*: 29–36.

Heuer, R., Kuhr, B., Fahimi, M., Curtin, T.R., Hinsdale, M., Carley-Baxter, L., and Green, P. (forthcoming). *National Study of Postsecondary Faculty (NSOPF:04) Methodology Report* (NCES 2005–xxx). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Iannacchione, V. (1982). *Weighted Sequential Hot Deck Imputation Macros*. Presented at the Seventh Annual SAS User's Group International Conference, San Francisco, February.

Iannacchione, V. (2003). Sequential weight adjustments for location and cooperation propensity for the 1995 National Survey of Family Growth. *Journal of Official Statistics*, *16*:31-43.

Kass, G.V. (1980). An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics, 29*: 119-127.

Kaufman, S. (2004). Using the Bootstrap in a Two-Stage Design when Some Second-Stage Strata have only One Unit Allocated. *Proceedings of the American Statistical Association, Section on Survey Research Methods*.

Folsom, R.E., and Singh, A.C. (2000). The Generalized Exponential Model for Sampling Weight Calibration for Extreme Values, Nonresponse, and Poststratification. *Proceedings of the Section on Survey Research Methods of the American Statistical Association,* 598–603.

Riccobono, J.A., Whitmore, R.W., Gabel, T.J., Traccarella, M.A., Pratt, D.J., and Berkner, L.K. (1997). *National Postsecondary Student Aid Study, 1995–96 (NPSAS:96), Methodology Report* (NCES 98–073). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

RTI International (RTI). (2004). *SUDAAN User's Manual, Release 9.0*. Research Triangle Park, NC: RTI International.

U.S. Department of Education. (1999). *NCES Restricted-Use Data Procedures Manual*. Washington, DC: National Center for Education Statistics.

U.S. Department of Education. (2003). *NCES Statistical Standards* (NCES 2003–601). Washington, DC: National Center for Education Statistics.

Woodruff, R.S. (1971). A Simple Method for Approximating the Variance of a Complicated Estimate. *Journal of the American Statistical Association*, *66*: 411–414.