National Postsecondary Student Aid Study (NPSAS)

Website: https://nces.ed.gov/surveys/npsas/
Updated: June 2018

SAMPLE SURVEY OF POSTSECONDARY INSTITUTIONS AND STUDENTS; CONDUCTED EVERY 3 to 4 YEARS

NPSAS collects information from:

- Student institutional record abstracts
- U.S. Department of Education administrative records
- Student interviews
- Parent interviews

1. OVERVIEW

National Postsecondary Student Aid Study (NPSAS) is a comprehensive nationwide study conducted every three to four years by the National Center for Education Statistics (NCES) to determine how students and their families pay for postsecondary education. It is designed to address policy questions resulting from the rapid growth of financial aid programs and the succession of changes in financial aid program policies since 1986. The first NPSAS was conducted in the 1986–87 academic year (NPSAS:87). The most recently completed in the series was administered in the 2015–16 academic year (NPSAS:16). Other administrations have been conducted in academic year 1989–90 (NPSAS:90), 1992–93 (NPSAS:93), 1995–96 (NPSAS:96), 1999–2000 (NPSAS:2000), 2003–04 (NPSAS:04), 2007–08 (NPSAS:08), and 2011–12 (NPSAS:12).

NPSAS is based on a nationally representative sample of all students in eligible postsecondary education institutions in the 50 states, the District of Columbia, and Puerto Rico[1]. Sampled institutions represent all major sectors, including public and private, nonprofit and for-profit, and less-than-2-year schools, community colleges, 4-year colleges, and major universities with graduate-level programs. Study members include both undergraduate and graduate students who receive financial aid as well as those who do not. NPSAS data are obtained from administrative records of student financial aid, interviews with students, and, in prior cycles, interviews with a subsample of parents.

NPSAS also provides baseline data for two longitudinal studies: the Beginning Postsecondary Students Longitudinal Study (BPS) and the Baccalaureate and Beyond Longitudinal Study (B&B; see respective chapters). NPSAS:90, NPSAS:96, NPSAS:04, and NPSAS:12 served as baselines for BPS cohorts; NPSAS:93, NPSAS:2000, NPSAS:08, and NPSAS:16 were the baselines for B&B cohorts.

Unlike prior administrations, 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)—was conducted primarily as a separate study, with the exception of institution sampling and contacting (see NSOPF chapter). NPSAS:04 and NPSAS:08 study samples were also supplemented to provide representative estimates by institutional sector for several states.

[1] Puerto Rico was not included in the 1987 and 2012 administrations of NPSAS.

Purpose
The purpose of NPSAS is to produce reliable national estimates of characteristics related to financial aid for postsecondary students, the role of financial aid in how students and their families finance postsecondary education, and the extent to which the financial aid system is meeting the needs of students and families.

Components
NPSAS collects data on students from several sources, including: student records at the institution attended, student interviews, the Federal Student Aid Central Processing System (CPS), the National Student Loan Data System (NSLDS), the National Student Clearinghouse (NSC), ACT and SAT files, the IPEDS Institutional Characteristics (IC) file, and the *Veterans Benefits Administration (VBA)*.

Student Record Collection. The following information on students is obtained from institutional records: year in school, major field of study, type and control of institution, attendance status, tuition and fees, admission test scores, financial aid awards, cost of attendance, student budget information and expected family contribution for aided students, grade point average, age, and date first enrolled. Typically, an appointed Institutional Coordinator or a field data collector extracts the information from student records at a sample institution and enters it into a secure, customized web data collection system. In some cases, institutions and centralized systems choose to create and transmit a data file containing this information for all sample students from the sample institution(s).

Student Interview. Web-based student interviews (completed as a telephone interview or by self-administration) provide data on level (undergraduate, graduate[2]), major field of study, financial aid at other schools attended during the year, other sources of financial support, education experiences, current marital status, age, race/ethnicity, sex, highest degree expected, and employment and income.

U.S. Department of Education Administrative Records. Since NPSAS:96, the following information has been collected from U.S. Department of Education Central Processing System (CPS) and National Student Loan Data System (NSLDS): types and amounts of federal financial aid received, cumulative Pell Grant and Direct loan amounts (also known as Stafford loan), and loan repayment status. In NPSAS:08, information was also obtained for recipients of the new Academic Competitiveness Grant (ACG) and the National Science and Mathematics Access to Retain Talent Grant (National SMART Grant). The ACG and National SMART grants were discontinued in the 2011–12 academic year, so information on these grants was not collected in NPSAS:12.

Other administrative databases. Data is also collected from commercial databases, such as: enrollment, degree, and certificate records from the National Student Clearinghouse (NSC); ACT and SAT test score data from ACT, Inc. and College Board. NPSAS:16 also included military service and veterans' education benefits information from the Veterans Benefits Administration (VBA).

Parent Interview. Telephone interviews with a limited sample of students' parents (conducted through NPSAS:96) collected supplemental data, including parents' marital status, age, highest level of education achieved, income, amount of financial support provided to children, types of financing used to pay children's educational expenses, and occupation and industry.

Out-of-School Student Loan Recipient Survey. This survey was only conducted as part of NPSAS:87. It collected data on major field of study; years attended and degrees received (if any); type and control of institution; financial aid; aid repayment status; age; sex; race/ethnicity; marital status; income; and employment history (occupation, industry, and salary).

Periodicity
Triennial from 1986–87 through 1995–96, and quadrennially beginning in 1999–2000. Beginning in 2017–18, an administrative data collection will also be conducted quadrennially. The next full data collection with a student interview is scheduled for 2019–20, and the first administrative data collection is scheduled for spring of 2018.

Data Availability
Public-use data from NPSAS:96 through NPSAS:16 is available at https://nces.ed.gov/datalab/index.aspx. Public-use data from NPSAS:87 through NPSAS:93 is available at https://nces.ed.gov/dasol/. Information on NPSAS restricted-use data files is available at https://nces.ed.gov/pubsearch/licenses.asp.

2. USES OF DATA

The goal of the NPSAS study is to identify institutional, student, and family characteristics related to participation in financial aid programs. Federal policymakers use NPSAS data to determine future federal policy concerning student financial aid. With these data, it is possible to analyze special population enrollments in postsecondary

[2] Previous administrations of NPSAS have included samples of first-professional students. However, IPEDS has replaced the term "first-professional" with "doctor's degree – professional practice".

education, including students with disabilities, racial and ethnic minorities, students taking remedial/developmental courses, students from families with low incomes, and older students. The distribution of students by major field of study can also be examined. Fields of particular interest are mathematics, science, and engineering, as well as teacher preparation and health studies. Data can also be generated on factors associated with choice of postsecondary institution, participation in postsecondary vocational education, parental support for postsecondary education, and occupational and educational aspirations.

It is important that statistical analyses be conducted using software that properly accounts for the complex sampling design of NPSAS. NCES developed software tools for analysis of complex survey data: for instance, QuickStats allows users to generate simple tables and graphs quickly, and PowerStats allows researchers to generate more complex tables and run linear and logistic regressions. Beginning with the NPSAS:96 collection, data can be analyzed with PowerStats. The Data Analysis System (DAS) may be used for analyses using NPSAS data prior to 1995–96. For information on other software packages and statistical strategies useful for analysis of complex survey data, see appendix K of the *2015–16 National Postsecondary Student Aid Study (NPSAS:16) Data File Documentation Report* (Wine et al. 2018).

3. KEY CONCEPTS

Described below are several key concepts relevant to financial assistance for postsecondary education. For additional NPSAS terms, refer to the glossaries in published statistical analysis reports and database documentation.

Level and Control of Institution. Control refers to the source of revenue and control of operations (public, private nonprofit, private for-profit), and level refers to the highest degree or award offered by the institution in any program. Doctorate-granting institutions award a doctoral—research/scholarship or doctoral—professional practice degree in one or more programs; non-doctorate-granting 4-year institutions award at least a bachelor's degree; 2-year institutions award at least an associate's degree; and less-than-2-year institutions award certificates or other credentials in vocational programs. Public 4-year, non-doctorate-granting institutions are further subdivided by whether the institutions primarily conferred subbaccalaureate awards (certificates and associate's degrees) or bachelor's degrees.

Attendance Pattern. A student's intensity and persistence of attendance during the NPSAS year. Intensity refers to whether the student attended full- or part-time while enrolled. Persistence refers to the number of months a student is enrolled during the year. Students are considered to be enrolled for a full year if they are enrolled 8 or more months during the year. Months do not have to be contiguous or at the same institution, and students do not have to be enrolled for a full month to be considered enrolled for that month.

Dependency Status. If a student is considered financially dependent, the parents' assets and income are considered in determining aid eligibility. If the student is financially independent, only the student's assets are considered, regardless of the relationship between student and parent. The federal definition of dependency status has remained the same in each administration of NPSAS since academic year 1995–96. All students who are age 24 or over in the fall term of the NPSAS year are considered to be independent. Students under 24 who are married, have legal dependents other than a spouse, are veterans, or are an orphan or ward of the courts are also independent. Other undergraduates under age 24 are considered to be dependent, unless they can demonstrate to a financial aid officer that they do not receive any financial support from their parents. All graduate students in programs beyond a bachelor's degree are considered to be independent.

Expected Family Contribution (EFC). The amount of financial support for the student's undergraduate education that is expected to be provided by the student's family, or directly by the student if the student is financially independent. This amount is used to determine financial need and is based upon dependency status (see above definition), family income and assets, family size, and the number of children in the family enrolled in postsecondary education. This information is gathered from the Department of Education's financial aid system (the Central Processing System), or it is imputed from student income.

Title IV Financial Aid. The sum of the following types of federal aid: Pell Grants, Supplemental Educational Opportunity Grants (SEOG), Perkins Loans, Direct Loans (also known as Stafford Loans), PLUS Loans, and Federal Work Study. NPSAS:08 also included Academic Competitiveness Grants and National SMART Grants.

4. SURVEY DESIGN

Target Population

The *target population* is defined as all eligible students enrolled at any time during the federal financial aid award year in postsecondary institutions in the United States or Puerto Rico[3] that have a signed Title IV participation

[3] Puerto Rico was not included in the 1987 and 2012 administrations of NPSAS.

agreement with the U.S. Department of Education (thus making these institutions eligible for federal student aid programs). The population includes both students who receive aid and those who do not receive aid. It excludes students who are enrolled solely in a general equivalency diploma (GED) program or are concurrently enrolled in high school.

Sample Design

The design for the NPSAS sample involves the selection of a nationally representative sample of postsecondary education institutions and students within these institutions. Prior to NPSAS:96, a geographic-area-clustered, three-stage sampling design was used to: (1) construct geographic areas from three-digit postal zip code areas; (2) sample institutions within the geographic sample areas; and (3) sample students within sample institutions. Beginning with NPSAS:96, the sample design eliminated the first stage of sampling (geographic area construction), thereby increasing the precision of the estimates. Institutional and student sample sizes vary somewhat from cycle to cycle depending on study design and budget considerations at the time. Approximately 2,000 institutions and 122,030 students were initially selected for participation in NPSAS:16.

Institution Sample. To be eligible for inclusion in the institution sample, an institution must satisfy the following conditions: (1) offer an education program designed for persons who have completed secondary education; (2) offer an academic, occupational, or vocational program of study lasting at least 3 months or 300 clock hours; (3) offer access to the general public; (4) offer more than just correspondence courses; (5) be located in the 50 states, the District of Columbia, or Puerto Rico[4]; and (6) be other than a U.S. Service Academy. Also, beginning with NPSAS:2000, eligible institutions must have a signed Title IV participation agreement with the U.S. Department of Education.

The institution-level sampling frame is constructed from the Integrated Postsecondary Education Data System (IPEDS) Institutional Characteristics (IC) and header files (see IPEDS chapter). Although the institutional sampling strata have varied across NPSAS administrations, in all years the strata are formed by classifying institutions according to control (public or private), level, and highest degree offering. The NPSAS:04 strata were also formed by Carnegie classification and state, and the NPSAS:08 strata were also formed by state. A stratified sample of institutions is then selected with probability proportional to size. School enrollment, as reported in the IPEDS, defines the measure of size; enrollment is imputed if missing in the IPEDS file. Institutions with expected frequencies of selection greater than unity are selected with certainty. The remainder of the institution sample is selected from the other institutions within each stratum. Although prior NPSAS administrations aggregated private for-profit 2-year and 4-year institutions into one sampling strata, NPSAS:12 split the two into separate strata to reflect the growth in enrollment in the for-profit sector.

Although prior NPSAS administrations aggregated public 4-year non-doctorate-granting institutions into one sampling strata, NPSAS:16 split it into two into separate strata: public 4-year institutions that were primarily subbaccalaureate and those that were primarily baccalaureate. The subbaccalaureate institutions were usually community colleges that predominantly awarded subbaccalaureate degrees while offering bachelor's degrees in only a small number of select fields. Splitting the public 4-year, non-doctorate-granting institutions into two strata, rather than combining them, allowed for oversampling and controlling the sample size of the subbaccalaureate institutions and students in them, including the baccalaureate recipients.

Additional implicit stratification is accomplished within each institutional stratum by sorting the stratum sampling frame in a serpentine manner. Implicit stratification allows the approximation of proportional representation of institutions on additional measures.

NPSAS:16 sampled 2,000 institutions, and serves as the base-year survey for the B&B:16 cohort of baccalaureate recipients. NPSAS:16 categorized institutions into 11 strata based on institution level and control. Within each institution stratum for NPSAS:16, additional implicit stratification was accomplished by the following classifications: (1) Historically Black Colleges and Universities (HBCUs) status; (2) Hispanic Serving Institutions (HSIs) status; (3) INSTCAT (institution category derived using the level of offerings reported on the IPEDS Institutional Characteristics component and the number and level of awards that were reported on the IPEDS Completions component); (4) Carnegie classifications of degree-granting postsecondary institutions; (5) the Office of Business Economics Region from the IPEDS Header file (Bureau of Economic Analysis of the U.S. Department of Commerce Region); (6) state and system for states with large systems (e.g., the SUNY and CUNY systems in New York, the state and technical colleges in Georgia, and the California State University and University of California systems in California); and (7) the institution measure of size. This implicit stratification helped ensure that the sample was approximately proportional to the population for these measures.

[4] Puerto Rico was not included in the 2012 administration of NPSAS.

NPSAS:12 sampled 1,690 of a universe of 7,050 institutions, and serves as the base-year survey the BPS:12 cohort of first-time beginning college students. NPSAS:12 categorized institutions into 10 strata based on institution level, control, and highest level of offering. Within each institution stratum, NPSAS statisticians accomplished additional implicit stratification by sorting the sampling frame within stratum by the following classifications: (1) historically Black colleges and universities (HBCU) indicator; (2) Hispanic-Serving Institutions (HSI) indicator; (3) Carnegie classifications of degree-granting postsecondary institutions; (4) 2-digit Classification of Instructional Programs code of the largest program for less-than-2-year institutions; (5) the Office of Business Economics Region from the IPEDS header file (Bureau of Economic Analysis of the U.S. Department of Commerce Region); (6) state and system for states with large systems, e.g., the SUNY and CUNY systems in New York, the state and technical colleges in Georgia, and the California State University and University of California systems in California; and (7) the institution measure of size.

In NPSAS:08, the implicit strata were formed using (1) Historically Black Colleges and Universities (HBCU) indicator; (2) Hispanic-Serving Institutions (HSI) indicator; (3) Carnegie classifications of postsecondary institutions; (4) the Office of Business Economics (OBE) Region from the IPEDS header file (Bureau of Economic Analysis of the U.S. Department of Commerce Region); and (5) an institution measure of size. Further implicit stratification was done for the State University of New York (SUNY) and City University of New York (CUNY) systems in New York, the state and technical colleges in Georgia, and the state universities in California.

The NPSAS:08 institution sampling frame was constructed from the 2004–05 IPEDS IC, header, and Fall Enrollment files and, because NPSAS:08 also serves as the base-year survey for a longitudinal cohort of baccalaureate recipients (i.e., B&B), the 2004–05 IPEDS Completions file. A total of 1,960 of the 6,780 institutions in the survey universe were selected for the NPSAS:08 sample. The sampled institutions were stratified into 22 national strata and 24 state strata based on institutional control, institutional offering, and highest degree offering.

In NPSAS:04, the implicit strata were formed using (1) the HBCU indicator; (2) Carnegie classifications (3) OBE Region; and (4) an institution measure of size. In NPSAS:2000, for less-than-2-year, 2-year, and private for-profit institutions, the implicit strata were formed using (1) institutional level of offering (where levels had been collapsed to form strata); (2) the OBE Region from the IPEDS header file; (3) the Federal Information Processing Standard (FIPS) state code; and (4) an institution measure of size. For public 4-year and private nonprofit 4-year institutions, the implicit strata were formed using (1) Carnegie classifications of institutions or groupings of Carnegie classifications; (2) the HBCU indicator; (3) the OBE Region from the IPEDS header file; and (4) an institution measure of size. In NPSAS:96, the implicit strata were formed using (1) institutional level of offering; (2) the IPEDS IC-listed U.S. Department of Commerce Region; and (3) an institution measure of size. Selected institutions are asked to verify their IPEDS classification (institutional control and highest level of offering) and the calendar system that they use (including dates that terms start).

The institutional sampling frame for NPSAS:04 was constructed from the 2000–01 IPEDS IC, header, and Fall Enrollment files; 1,670 of the 6,706 institutions in the survey universe were selected for NPSAS:04. The sampled institutions were stratified into 22 national strata and 36 state strata based on institutional control, institutional offering, highest degree offering, and Carnegie classification. The institutional sampling frame for NPSAS:2000 was constructed from the 1998–99 IPEDS IC file and, because NPSAS:2000 also served as the base-year survey for a B&B cohort, the 1996–97 IPEDS Completions file. Eligible institutions were partitioned into 22 institutional strata based on institutional control, highest level of offering, and percentage of baccalaureate degrees awarded in education. Approximately 1,100 institutions were initially selected for NPSAS:2000. As noted above, NPSAS:96 was the first administration of NPSAS to employ a single-stage institutional sampling design, no longer constructing geographic areas as the initial step.

Student Sample. Full- and part-time students enrolled in academic or vocational courses or programs at eligible institutions, and not concurrently enrolled in a high school completion program, are eligible for inclusion in NPSAS. NPSAS:87 sampled students enrolled in the fall of 1986. Beginning with NPSAS:90, students enrolled at any time during the year were eligible for the study. This design change provided the data necessary to estimate full-year financial aid awards.

Sampled institutions are asked to provide student enrollment lists with the following information for each student: full name, identification number, Social Security number, educational level, an indication of first-time beginning student (FTB) status or baccalaureate recipiency (depending on the longitudinal cohort being launched), major, and, beginning with NPSAS:04, a local address, a local telephone number, a campus e-mail, a permanent address, a permanent phone number, and a permanent e-mail. Additionally, beginning with NPSAS:08, date of birth and class level of undergraduates are requested. The student sample is drawn from the enrollment lists, which were provided by 1,750 of 1,990 eligible institutions for

NPSAS:16; 1,480 of 1,690 eligible institutions for NPSAS:12; 1,730 of 1,940 eligible institutions in NPSAS:08; 1,360 of 1,630 eligible institutions in NPSAS:04; 1,000 of the nearly 1,100 eligible institutions in NPSAS:2000; and 840 of 900 eligible institutions in NPSAS:96.

Basic student sample. Students are sampled on a flow basis (using stratified systematic sampling) from the lists provided by institutions. Steps are taken to eliminate both within- and cross-institution duplication of students. NPSAS classifies students by educational level as undergraduate, master's, doctor's, other graduate, or professional students. NPSAS:16 further classified students into 17 total strata based on program type, veteran status, and whether the student was a baccalaureate recipient. For the purpose of defining the third cohort of B&B, NPSAS:08 classified undergraduates into (1) business major potential baccalaureate recipients, (2) other potential baccalaureate recipients, and (3) other undergraduates. Potential baccalaureate recipients were further stratified by those who are science, technology, engineering, or mathematics (STEM) majors and all other majors and by SMART Grant recipients and non-recipients. Other undergraduates were further stratified by SMART Grant recipients, Academic Competitiveness Grant (ACG) recipients, and non-recipients. The categories for potential baccalaureate recipients and other undergraduates were then stratified by in-state and out-of-state status. NPSAS:04 stratified undergraduate students as (1) potential FTBs and (2) other undergraduates. These two categories were then stratified by in-state and out-of-state status. The FTBs in NPSAS:04 make up the third cohort of BPS. For the purpose of defining the second cohort of B&B, NPSAS:2000 also broke down undergraduate categories into: (1) business major baccalaureate recipients, (2) other baccalaureate recipients, and (3) other undergraduates. In NPSAS:96, FTBs, or students beginning their postsecondary education during one of the terms of the NPSAS:96 sample year composed the second cohort of the BPS, with the data collected serving as the base-year data for the subsequent longitudinal studies.

The student sample is allocated to the combined institutional and student strata (e.g., graduate students in public 4-year doctorate institutions). Initial student sampling rates are calculated for each sample institution using refined overall rates to approximate equal probabilities of selection within the institution-by-student sampling strata. These rates are sometimes modified to ensure that the desired student sample sizes are achieved.

For NPSAS:16, initial student sampling rates were calculated for each sample institution using sampling rates designed to generate approximately equal probabilities of selection within the institution-by-student sampling strata. In certain instances, NPSAS statisticians modified sampling rates as follows: 1) Student sampling rates were increased for each institution to yield at least 10 students (if possible) to ensure sufficient yield for variance estimation. 2) Student sampling rates were decreased, with few exceptions, if an institution sample size was greater than 300 students. 3) Student sampling rates were adjusted higher or lower based on expected yield calculations for institutions where the sample had not yet been selected. These adjustments to the initial sampling rates resulted in some additional variability in the student sampling rates and increased survey design effects. For NPSAS:16, the expected sample of students was 126,320, with 122,030 being achieved, of which 37,890 students were potential baccalaureate recipients and 22,950 students were graduate level.

In NPSAS:12, adjustments were also made to the initial sampling rates. For NPSAS:12, the targeted sample of students was 124,650, with 128,120 being achieved, of which 59,740 students were undergraduate FTBs and 17,330 students were graduate level.

Initial sampling rates were adjusted in NPSAS:08, NPSAS:04, NPSAS:2000, and NPSAS:96, as well. The overall sample yield in NPSAS:08 was close to expected (137,800 students vs. the target of 138,000). The student sample consisted of 29,470 potential baccalaureate recipients; 95,650 other undergraduates; 6,530 master's students; 3,760 doctoral students; 470 other graduate students; and 1,920 first-professional students. The overall sample yield in NPSAS:04 was less than expected (109,210 students vs. the target of 121,680). The student sample consisted of 49,410 FTBs; 47,680 other undergraduates; 3,720 master's students; 4,950 doctoral students; 1,660 other graduate students; and 1,790 first-professional students. (See "FTB sample" below for more detail on the sampling of FTBs.) In NPSAS:2000, the overall sample yield was very close to expected (70,230 students vs. the target of 70,270). The student sample consisted of 57,600 undergraduates; 5,960 master's students; 3,950 doctoral students; 1,370 other graduate students; and 1,350 first-professional students. In NPSAS:96, the overall sample yield was actually greater than expected (63,620 students vs. the target of 59,510). The student sample consisted of 23,610 potential FTBs; 27,540 other undergraduates; 9,690 graduate students; and 2,780 first-professional students.

Student interview sample. NPSAS:04 was the first administration of NPSAS to offer the option of self-administration of the student interview via the Web, in addition to computer-assisted telephone interviewing (CATI). In NPSAS:16, there were about 77,030 completed full interviews (excludes 1,830 partial completes) with

39,020 completed by self-administration via web without telephone contact; 26,450 completed via self-administration via web with telephone contact; and the remaining 11,570 being completed via CATI. In NPSAS:12, there were approximately 85,000 completed interviews with 36,770 completed by self-administration (web without telephone contact); 31,710 completed via the web with telephone contact; and the remaining 14,820 being completed via telephone. In NPSAS:08, these procedures resulted in 95,360 completed interviews, about two-thirds of which were completed by self-administration and one-third by CATI. In NPSAS:04, these procedures resulted in 62,220 completed interviews, 28,710 of which were completed by self-administration and 33,510 by CATI.

In NPSAS:2000, student interviews were conducted primarily by CATI. To help reduce the level of nonresponse to CATI, computer-assisted personal interviewing (CAPI) procedures, using field interviewers, were used for the first time. Of the 66,340 eligible students in the initial CATI sample, some 51,010 were located for CATI interviewing, while 11,960 were "unlocatable" in CATI and were eligible for field locating and/or CAPI; the rest were either ineligible or excluded.

Due to budget limitations, NPSAS:96 attempted CATI interviews for only a subsample of the basic student sample. A two-phase, nonrespondent follow-up subsampling design was used to maximize the yield of completed student interviews obtained from the CATI subsample while achieving acceptable response rates. These procedures resulted in 51,200 students being selected for Phase 1 of the CATI interviewing. A sample of nonrespondents to Phase 1 was selected for Phase 2 with specified rates based on the outcome of the Phase 1 efforts and the seven sampling strata; 25,770 students were selected for Phase 2.

Parent interview subsample. In NPSAS:96, a subsample of students selected for the student interview was also designated for parent interviews. In the Phase 1 CATI subsample of NPSAS:96, students were designated for parent interviews if they met one of the following criteria: they were dependent undergraduate students not receiving federal aid; they were dependent undergraduate students receiving federal aid whose parents' adjusted gross income was not available; or they were independent undergraduate students who were 24 or 25 years old on December 31, 1995. All 8,800 students who fell into one of these groups were sampled for parent interviews. The parent interview was discontinued after NPSAS:96.

Longitudinal Study Samples. In NPSAS:90, a new longitudinal component collected baseline data for students who started their postsecondary education in the 1989–90 academic year. These students were followed over time in BPS, with the first follow-up in 1992. Beginning postsecondary students from NPSAS:96, NPSAS:04, and NPSAS:12 were also followed up and surveyed two and five years later. Similarly, NPSAS:93, NPSAS:2000, NPSAS:08, and NPSAS:16 provided baseline data for students who received baccalaureates in the 1992–93, 1999–2000, 2007–08, and 2015–16 academic years, respectively. These graduates have been followed over time as part of B&B.

BPS sample. To be eligible for BPS, students must have begun their postsecondary education for the first time, after completing high school, on or after July 1. NPSAS survey staff pay particular attention to accurately identifying FTBs in NPSAS to avoid unacceptably high rates of misclassification that were observed in past BPS studies, particularly false positives. High rates of misclassification can, and have, resulted in (1) excessive cohort loss, (2) excessive cost to "replenish" the sample, and (3) an inefficient sample design (excessive oversampling of "potential" FTBs) to compensate for anticipated misclassification error.

The participating institutions and several administrative data sources provided data to aid in properly classifying FTBs. Key data the institutions provided included an FTB indicator, high school graduation date, and date of birth. Administrative data sources, including the NSLDS, CPS, and National Student Clearinghouse (NSC), provided data that was of particular use in identifying false positives. Of the 719,450 students that the NPSAS staff sent to NSC for the NPSAS:12 data collection, about 7 percent were false positives.

B&B sample. The first B&B longitudinal cohort was identified in NPSAS:93 and consisted of students who received their bachelor's degree in academic year 1992–93. NPSAS:93 provided the base-year data, and students were interviewed in an initial follow-up in 1994; this follow-up also included a collection of transcript data. The 1993 cohort was surveyed again in 1997 and 2003. The first transcript collection was conducted as part of B&B:93/94. The second B&B cohort was selected from NPSAS:2000, which became the base year for a single follow-up in spring 2001.

B&B:08 was the third cohort in the B&B series and the second to gather college transcript data on such a longitudinal sample. The B&B:08 sample consisted of students eligible to participate in the NPSAS:08 full-scale study who completed requirements for the bachelor's degree in the 2007–08 academic year. The first follow-up study (B&B:08/09) involved two data collection components. First, postsecondary transcripts were collected from each of the NPSAS institutions where

sample members completed their program requirements. It was followed by an interview focusing on plans after degree completion.

B&B:16 is the fourth cohort in the B&B series. The B&B:16 sample consisted of students eligible to participate in the NPSAS:16 full-scale study who completed requirements for a bachelor's degree in the 2016–17 academic year. NCES contacted the B&B:16 cohort in 2017 for a follow-up survey and plans to do so again in 2020 and 2026.

B&B status is determined on the basis of multiple sources: student enrollment lists from institutions, student record collection, student interviewing, and transcripts (in B&B:93/94 and B&B:08/09).

Data Collection and Processing

Reference Dates. Data are collected for the financial aid award year, which spans from July 1 of one year through June 30 of the following year.

Data Collection. NPSAS involves a multistage effort to collect information related to student aid. The first stage involves collecting applicants from the U.S. Department of Education's Central Processing System (CPS).

Another stage of data collection involves collecting information from the student's records at the school from which he or she was sampled. Since NPSAS:93, these data have been collected through a computerized system, which facilitates both the collection and transfer of information to subsequent electronic systems. To reduce respondent burden, several data elements are preloaded into the records collection system records prior to collection at the institution. These include student demographics, Student Aid Report (SAR) information on federal financial aid applicants, and nonfederal aid common to a particular institution. Institutional Coordinators are given the option of having their staff or contractor field data collectors perform the data collections. About 66 percent of the institutions in NPSAS:04, as well as 74 percent in NPSAS:2000, and 57 percent in NPSAS:96 chose self-administration, using a computer-based program to provide student record data. In NPSAS:08, very few institutions (about 1 percent) chose the field interviewer option for completion. Approximately 63 percent chose self-administration, and 36 percent provided the student record data via electronic files (primarily large institutions or systems).

NPSAS:12 used four modes for student record abstraction: (1) Case Mode, in which institution staff entered data directly into the web-based system one student at a time, either by section or by student; (2) Grid Mode, in which institution staff entered data directly into the web-based system for multiple students at a time in a format resembling a grid; (3) Template Upload, in which institution staff downloaded an Excel template, entered data into it, then uploaded it back to the website; and (4) Data Files Upload, in which institution staff created data files following provided specifications. For NPSAS:12, 39 percent of the institutions keyed data into the web-based student record application via Case Mode or Grid Mode, 37 percent uploaded the Excel Template, and about 24 percent used the Data File Upload.

For NPSAS:16, student records could be completed in three modes via the Postsecondary Data Portal (PDP), a web-based student records interface: (1) Web mode, in which institution staff used drop-down boxes and text-entry fields to key data directly on the PDP website, one student at a time; (2) Excel mode, in which institutions downloaded a preformatted Excel spreadsheet template from the PDP, keyed or copied student data into a spreadsheet template offline, and then uploaded the completed template to the PDP website; and (3) Comma-separated values (CSV) mode, in which institutions downloaded customized file specifications from the PDP website, prepared data files offline according to the file specifications, and then uploaded completed files to the PDP website. In NPSAS:16, most institutions opted for the Excel mode (62 percent), 30 percent uploaded a CSV file, and the remaining 8 percent used the Web mode and entered data directly into the PDP student records interface.

In the student interview stage of data collection, information on family characteristics, demographic characteristics, and educational and work experiences and aspirations is obtained from students. Student and parent paper questionnaires were used to collect this information in NPSAS:87, but beginning with NPSAS:90, student and parent data were collected by computer-assisted-telephone-interviewing (CATI). Parent interviews, however, were not conducted after NPSAS:96. NPSAS:04 was the first administration of NPSAS to offer students the opportunity to participate by self-administered web surveys or by CATI, an approach that has continued in subsequent NPSAS administrations (i.e., NPSAS:08, NPSAS:12, and NPSAS:16).

The NPSAS:08 student interview contained seven sections and was programmed for both self-administered web surveys and CATI. 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. The abbreviated interview was also translated into Spanish for telephone administration to Spanish speakers with limited English proficiency.

The student interview included an online coding system used to obtain IPEDS information for postsecondary institutions (other than the NPSAS institution from which the student was sampled) that the student attended during the same year. After the respondent or interviewer provided 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/interviewer could select the appropriate institution. Upon selection, the name of the institution, as well as selected IPEDS variables (institutional level, control), was inserted into the database.

An assisted coding system was also developed to facilitate the coding of major/field of study into categories that can be mapped to values in NCES's Classification of Instructional Programs (CIP).

The data collection design for student interviewers has evolved over time. In NPSAS:2000, student interviews were conducted primarily by telephone, and occasionally in person, using CATI/CAPI technology. In NPSAS:04 and NPSAS:08 abbreviated interviews were developed to convert refusals toward the end of data collection, and an online coding system was used, to obtain IPEDS information. NPSAS:96 differed from other cycles in that only a subsample of the initial student sample was selected for the interview stage (in order to reduce overall costs for the study).

The final stage of data collection involves retrieval of additional SAR data (for the academic year beyond the NPSAS year) from the Central Processing System (CPS), data on Pell Grant applications for the NPSAS year from the Pell Grant file, and data on recipients of Academic Competitiveness Grants and SMART Grant, as well as loan histories of applicants for federal student loans from the National Student Loan Data System (NSLDS). All of these files are maintained by the U.S. Department of Education. Additional data for the NPSAS sample are obtained from other sources as well, including test score data from the ACT and College Board (SAT), enrollment data from the National Student Clearinghouse (NSC), and data from the Veterans Benefits Administration (VBA).

Editing. Initial editing takes place during data entry. The web-based data collection systems used for the student interview and student record collection have built-in quality control checks to notify users of invalid or out-of-range entries. For example, the student records collection system will notify the user of any student records that are incomplete (and the area of incompleteness) and any records that have not yet been accessed. A pop-up screen provides full and partial completion rates for institutional record collection. Data are subjected to edit checks for completeness of critical items.

Following the completion of data collection, all student record and interview data are edited to ensure adherence to range and consistency checks. Range checks are summarized in the variable descriptions contained in the data files. Inconsistencies, either between or within data sources, are resolved in the construction of derived variables. Items are checked for validity by comparing the student interview responses to information available in institutional records. Missing data codes characterize blank fields as don't know/data not available; refused; legitimate skip; data source not available (not applicable to the student); or other.

Estimation Methods

Weighting is used to adjust NPSAS data to national population totals and to adjust for unit nonresponse. Imputation is used to compensate for item nonresponse and mitigate associated bias.

Weighting. For the purpose of obtaining nationally representative estimates, sample weights are created for both the institution and the student. Additional weighting adjustments, including nonresponse and poststratification adjustments, compensate for potential nonresponse bias and frame errors (differences between the survey population and the ideal target population). The weights are also adjusted for multiplicity at the institution and student levels and for unknown student eligibility.

In NPSAS:04 through NPSAS:16, the institution weight was computed first and then used as a component of the student weight. Student weights were calculated as the product of the total of 9 weight components for NPSAS:16; NPSAS:12 used 12 weight components, NPSAS:08 used 10 weight components, and NPSAS:04 used 13 weight components. Each represented either a probability of selection or a weight adjustment.

In NPSAS:2000, statistical analysis weights were computed for two sets of respondents: CATI respondents and other study respondents. These were calculated as the product of 13 weight components, again representing either a probability of selection or a weight adjustment.

In NPSAS:96, study weights were applied to students who responded to specified student record or CATI data items. Study and CATI weights were calculated as the product of 14 weight components. First-time beginning students (FTBs) whose first postsecondary institution was not the NPSAS sample institution were not included in BPS. To compensate for their exclusion, FTB weights were computed by making a final weighting class adjustment to the CATI weights by institution type. All adjustment factors were close to one, ranging from 1.00 to 1.02. The development of the student record weight components was similar to the development of the study and CATI weight components—except that the student record components

applied to a different set of respondent data and did not include the CATI weight components.

Imputation. When the editing process (including logical imputations) is complete, the remaining missing values for all variables with missing data are statistically imputed in order to reduce the bias of survey estimates caused by missing data. Variables are imputed using a weighted sequential hot-deck procedure whereby missing data are replaced with valid data from donor records that match the recipients with respect to the matching criteria.

For NPSAS:16, missing data were imputed for all variables. The imputation procedures involved a four-step process. In the first step, missing values were logically imputed. In the second step, variables and groups of variables were prioritized for imputation based upon their level of missing data; those with low levels of missingness were imputed before those with greater levels of missingness. In the third step, an initial weighted sequential hot deck (WSHD) process was implemented. Finally, in the fourth step, a cyclic *n*-partition hot deck process was implemented to iteratively cycle through *n*-partition hot decks. For NPSAS:12, missing data were imputed for all variables included in the restricted-use derived file. After replacing missing data in those cases where values could be deduced with certainty based upon logical relationships among observed variables, the weighted sequential hot deck (WSHD) method was used to replace missing data by imputing plausible values from statistically selected donor cases.

In NPSAS:08 and NPSAS:04, 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 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 the variables. For variables with more than 5 percent missing data, a statistical process called Chi-Squared Automatic Interaction Detection (CHAID) was used to identify the matching criteria that were most closely related to the variables being imputed.

In NPSAS:2000, the remaining missing values for 23 analysis variables were imputed statistically; most of the variables were imputed using a weighted hot-deck procedure. To implement the weighted hot-deck procedure, imputation classes and sorting variables relevant to each item being imputed were defined. If more than one sorting variable was chosen, a serpentine sort was performed where the direction of the sort (ascending or descending) changed each time the value of a variable changed. The serpentine sort minimized the change in the student characteristics every time one of the variables changed its value.

The respondent data for five of the items being imputed were modeled using a CHAID analysis to determine the imputation classes. These items were parent income (imputed for dependent students only), student income (imputed for independent students only), student marital status, local residence, and a dependents indicator.

A CHAID analysis was performed on these variables because of their importance to the study and the large number of candidate variables available with which to form imputation classes. Also, for the income variables, trying to define the best possible imputation classes was important due to the large amount of missing data. The CHAID analysis divided the respondent data for each of these five items into segments that differed with respect to the item being imputed. The segmentation process first divided the data into groups based on categories of the most significant predictor of the item being imputed. It then split each of these groups into smaller subgroups based on other predictor variables. It also merged categories of a variable that were found insignificant. This splitting and merging process continued until no more statistically significant predictors were found (or until some other stopping rule was met). The imputation classes were then defined from the final CHAID segments.

In NPSAS:96, some 22 analysis variables were statistically imputed. All variables, with the exception of the estimated family contribution were imputed using a weighted hot-deck procedure. First, the respondent data for six key items were modeled using a CHAID analysis to determine the imputation classes. These items were race/ethnicity, parent income (for dependent students only), student income, student marital status, a dependents indicator, and number of dependents. Then, 21 items imputed by the weighted hot-deck approach. The remaining 15 items were: parent family size, parent marital status, student citizenship, student gender, student age, dependency status, local residence, type of high school degree, high school graduation year, fall enrollment indicator, attendance intensity in fall term, student level in last term, student level in first term, degree program in last term, and degree program in first term. Only four of these 15 items had

more than 5 percent of their cases imputed: parent family size (18 percent), parent marital status (16 percent), high school degree (5 percent), and high school graduation year (5 percent).

Recent Changes

In prior NPSAS administrations, federal student loans older than 10 years as of the beginning of the study were excluded from cumulative borrowing and outstanding loan amount variables. In NPSAS:16, this was changed so that loans older than 10 years were included in these variables. As a result, cumulative borrowing estimates in NPSAS:16, especially for older student subpopulations, may differ from estimates for prior NPSAS administrations, with prior studies underestimating these amounts.

Prior to NPSAS:16, certain state grants that were administered by institutions (similar to how federal campus-based aid is administered) were classified as institutional grants. Because federal campus-based aid programs are classified by the source of funds and not by who administers the aid, this practice was changed in NPSAS:16, where campus-based state grants are now classified as state aid. This change in methodology mainly affects the aid of undergraduate students attending public institutions in California but led to larger population estimates of state grants and smaller estimates of institutional grants in NPSAS:16 compared with past NPSAS studies. To promote the analysis of trends in state and institutional aid over time, new state and institutional aid variables were added to the undergraduate files of NPSAS:96, NPSAS:2000, NPSAS:04, NPSAS:08, and NPSAS:12. These variables remove campus-based state grants from institutional grants and add them to state grant variables to be comparable with the NPSAS:16 methodology.

In NPSAS:16, an administrative data match to VBA databases was conducted to obtain information on sampled students' receipt of federal veterans' education benefits and their military service. The VBA data was the sole source for federal veterans' education benefits amounts, and they include payments for tuition and fees, books and supplies, work-study, housing, and other education expenses. Estimates of federal veterans' education benefits in prior NPSAS cycles were derived from self-reported amounts, amounts reported by the recipient's NPSAS institution, and stochastic imputation and were significantly lower on average than amounts in NPSAS:16. These earlier values may not include all the benefits included in the VBA data, particularly housing benefits, which were not explicitly requested from students or their institutions.

For NPSAS:12 and NPSAS:16, sample members were classified as a *study member* if data were available for him or her on a set of key variables, and these study members are the unit of analysis for those collections.

Changes to the NPSAS:12 student interview included core data elements used in previous NPSAS student interviews as well as new data elements developed in association with a redesign of the BPS longitudinal follow-up study. Additionally, 20 newly eligible institutions were included in the sample, using newly available 2009–10 IPEDS IC header, 12-Month and Fall Enrollment, and Completions components to create an updated sampling frame of current NPSAS-eligible institutions.[5]

NPSAS:04 included important new features in sample design and data collection. For the 2004 study, NPSAS and NSOPF were conducted together under one contract: the 2004 National Study of Faculty and Students (NSoFaS:04). There has historically been a great deal of overlap in the institution samples for these two studies since the target populations for both involve postsecondary institutions. To minimize institutional burden, and to maximize efficiency in data collection procedures, the two studies were combined.

Another important change in NPSAS:04 was that it 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 nonprofit 4-year institutions—in 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. NPSAS:08 was designed to provide state-level representative estimates for undergraduates within four institutional strata—public 2-year institutions, public 4-year institutions, private nonprofit 4-year institutions, and private for-profit degree-granting 2-year-or-more institutions. In NPSAS:08, state-level estimates were provided for California, Texas, New York, Illinois, Georgia, and Minnesota.

Also of importance is the inclusion of an option for self-administration via the Web of the student interview in NPSAS:04. This option was provided in addition to CATI interviews, which were employed in past rounds of NPSAS. Regardless of completion mode, a single web-based instrument was employed.

NPSAS:08 was again conducted independently of the NSOPF study but carried along all of the technical innovations and design enhancements of prior rounds. It was also designed to provide state-level representative

[5] Puerto Rico was not included in the 2012 administration of NPSAS.

estimates for undergraduates within four institutional strata—public 2-year institutions, public 4-year institutions, private nonprofit 4-year institutions, and private for-profit degree-granting 2-year-or-more institutions. In NPSAS:08, state-level estimates were provided for California, Texas, New York, Illinois, Georgia, and Minnesota.

The most significant enhancement to NPSAS:2000 involved the development and implementation of a new web-based system for use in the student record abstraction process. This web-based software had an improved user interface compared to the NPSAS:96 system and addressed several of the student records collection issues raised during NPSAS:96 (e.g., insufficient computer memory, failures during diskette installation and virus scanning, and lack of information regarding institutions' progress during data collection).

Other changes in NPSAS:2000 included: adding a series of questions about financial aid, as a new way of obtaining information about financial assistance received from sources other than federal student aid; adding several new items intended to capture the increased use of technology among students; and adding a new eligibility requirement for postsecondary institutions—to have a signed Title IV participation agreement with the U.S. Department of Education during the NPSAS academic year.

NPSAS:96 introduced important new features in sample design and data collection. It was the first NPSAS to employ a single-stage institutional sampling design (no longer using an initial sample of geographic areas and institutions within geographic areas). This design change increased the precision of study estimates. NPSAS:96 was also the only NPSAS to select a subsample of students for telephone interviews and to take full advantage of administrative data files. Through file matching/downloading arrangements with the Department of Education's Central Processing System, the study obtained financial data on federal aid applicants for both the NPSAS year and the following year. Through similar arrangements with the National Student Loan Data System, full loan histories were obtained. Cost efficiencies were introduced through a dynamic two-phase sampling of students for CATI, and the quality of collected institutional data was improved through an enhanced student records collection procedure. New procedures were also introduced to broaden the base of postsecondary student types for whom telephone interview data could be collected: the use of Telephone Display for the Deaf technology to facilitate telephone communications with hearing-impaired students, and a separate Spanish translation interview for administration to students with limited English language proficiency.

Future Plans

The next NPSAS data collection (NPSAS:18-AC) is scheduled for the 2017–18 academic year. Future NPSAS collections will continue to include a student interview every four years (NPSAS:16, NPSAS:20, NPSAS:24) to yield nationally representative data. In alternating cycles, an Administrative Collection (NPSAS:18-AC, NPSAS:22-AC, and NPSAS:26-AC) will be conducted in which only administrative data from the Department's data systems and institutional student records will be compiled to yield state-representative data.

5. DATA QUALITY AND COMPARABILITY

Every major component of the study is evaluated on an ongoing basis so that necessary changes can be made and assessed prior to task completion. Separate training is provided for Student Records (previously referred to as CADE) and CATI data collectors, and interviewers are monitored during CATI operations for deviations from item wording and skipping of questions. The CATI system includes online coding of postsecondary education institution and major field of study, so that interviewers can request clarification or additional information at the time of the interview. Quality circle meetings of interviewers, monitors, and supervisors provide a forum to address work quality, identify problems, and share ideas for improving operations and study outcomes. Even with such efforts, however, NPSAS—like every survey—is subject to various types of errors, as described below.

Sampling Error

Because NPSAS samples are probability-based samples rather than simple random samples, simple random sample techniques for estimating sampling error cannot be applied to these data. Two procedures for estimating variances, the Taylor Series linearization procedure and the Jackknife replicate procedure, are available for use with NPSAS:96 data. The Taylor Series linearization procedure and the balanced repeated replication (BRR) procedure are available on the NPSAS:2000 data files. The Taylor Series linearization procedure and the bootstrap replication procedure are available on the NPSAS:12, NPSAS:08, and NPSAS:04 data files.

Taylor Series. For NPSAS:96, analysis strata and replicates for three separate datasets were defined: all students, all undergraduate students, and all graduate/first-professional students. For NPSAS:2000, analysis strata and replicates for four separate datasets were defined: all students, all undergraduate students, all graduate/first-professional students, and all baccalaureate recipients. Beginning with NPSAS:04, analysis strata and replicates were defined for the combined set of all students.

Jackknife. In NPSAS:96, the Jackknife analysis strata were defined to be the same as the analysis strata defined for the Taylor Series procedure. Based on the Jackknife strata and replicate definitions, seven replicate weight sets were created—one set for the CADE weights and three sets each for the study and CATI weights. The study and CATI sets included separate replicate weights for all students, undergraduates only, and graduates only.

Balanced Repeated Replication. The BRR procedure is an alternative variance estimation procedure that computes the variance based on a balanced set of pseudo-replicates. To form pseudo-replicates for BRR variance estimation, the Taylor Series analysis strata were collapsed. The number of Taylor Series analysis strata and primary sampling units were different for all students combined, graduates/first-professionals, and baccalaureate recipients, so the collapsing was done independently and, hence, with different results. Replicate weights were created, associated with the two analysis weights: study weights and CATI weights. Thus, a total of five replicate weight sets were created for NPSAS:2000. For the study weights, this included separate replicate weights for all students and for graduate/first-professional students only; for the CATI weights, this included separate replicate weights for all students, graduate/first-professional students only, and baccalaureates only.

Bootstrap. In NPSAS:08 and NPSAS:04, a vector of bootstrap sample weights was added to the analysis file to facilitate computation of standard errors for both linear and nonlinear statistics. 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 purpose of computing the desired estimates. The vector of replicate weights allows for computing additional estimates for the sole purpose of estimating a variance. The replicates in NPSAS:16, NPSAS:12, and NPSAS:08 were produced using methodology adapted from Kott (1998) and Flyer (1987) and those in NPSAS:04 were produced using a methodology and computer software developed by Kaufman (2004). NPSAS:16, NPSAS:12, and NPSAS:08 included 200 replicate weights.

Nonsampling Error

Coverage error. Because the institutional sampling frame is constructed from the IPEDS IC file, there is nearly complete coverage of the institutions in the target population. Student coverage, however, is dependent upon the enrollment lists provided by the institutions. In NPSAS:16, approximately 1,750 of the 1,990 eligible institutions provided enrollment lists. For NPSAS:12, approximately 1,480 of the 1,690 eligible institutions provided enrollment lists. In NPSAS:08, approximately 1,730 of the 1,940 eligible institutions provided student lists or databases that could be used for sample selection. A total of 1,360 of the 1,630 eligible institutions in NPSAS:04; 1,000 of the nearly 1,100 eligible institutions in NPSAS:2000; and 840 of the 900 eligible institutions in NPSAS:96 provided student lists or databases that could be used for sample selection.

Several checks for quality and completeness of student lists are made prior to actual student sampling. In NPSAS:96 and NPSAS:04, *completeness checks* failed if (1) FTBs were not identified (unless the institution explicitly indicated that no such students existed) or (2) student level (undergraduate, graduate, or first professional) was not clearly identified. In NPSAS:2000 and NPSAS:08, completeness checks failed if (1) baccalaureate recipients/graduating seniors were not identified, (2) student level was not clearly identified, or (3) major fields of study or CIP codes were not clearly identified for baccalaureates.

Quality checks are performed by comparing the unduplicated counts (by student level) in institution lists with the nonimputed unduplicated counts in IPEDS IC files. Institutions failing these checks were called to rectify the problems before sampling began. These checks were performed through the 2007–08 administration. In NPSAS:08, after any necessary revisions, all but seven lists submitted were usable for selecting the student sample; in NPSAS:04, all but two lists submitted were usable for selecting the student sample.

For NPSAS:12, institutions were contacted if quality and completeness checks failed for the requested list of data items, which included: student's name; Social Security number; student ID number (if different than Social Security number); student level (undergraduate, masters, doctoral-research/scholarship/other, doctoral-professional practice, other graduate); FTB indicator; class level of undergraduates (first year, second year, etc.); date of birth; Classification of Instructional Program code or major; undergraduate degree program; high school graduation date (month and year); and contact information (local and permanent street address and telephone number and school and home e-mail address).

For NPSAS:16, once institutions submitted enrollment lists, NPSAS project staff performed several checks on the quality and completeness before selecting the student sample. These included verifying that institutions used a readable format and that key data needed for sampling and initial locating (e.g., baccalaureate indicator, SSN, contact information) were provided. If staff detected problems with lists during quality checks, they contacted institutions to resolve any issues.

Nonresponse error. *Unit nonresponse.* For NPSAS:16, there were 1,750 respondent institutions from among the

1,990 eligible sample institutions (88 percent unweighted and 90 percent weighted). The institution-weighted response rate was below 85 percent for 3 of the 11 types of institutions: public less-than-2-year; private for-profit, less-than-2-year; and private for-profit, 2-year institutions. The weighted response rates, by control and level of institution, ranged from 74 percent for private for-profit, less-than-2-year institutions to 95 percent for public 4-year, nondoctorate-granting, primarily subbaccalaureate institutions. Table NPSAS-1 provides a summary of response rates across NPSAS administrations.

For NPSAS:12, there were 1,480 respondent institutions from among the 1,690 eligible sample institutions (88 percent unweighted and 87 percent weighted). The institution weighted response rate was less than 85 percent for five of the ten institution types: public less-than-2-year institutions; public 2-year institutions; private, nonprofit, less-than-4-year institutions; private, for-profit, less-than-2-year institutions; and private for-profit 2-year institutions. The weighted response rates, by type of institution, ranged from 78 percent for private nonprofit less-than-4-year institutions and private for-profit 2-year institutions to 92 percent for public 4-year non-doctorate-granting institutions. Because study members, not interview respondents, were the unit of analysis in NPSAS:12, only a study member weight was created. As a result, analysts could not compare nonresponse bias analyses after weight adjustments. For more information, see Wine, Bryan, and Siegel, 2014.

For NPSAS:08, some 90 percent (weighted) of eligible sample institutions provided student enrollment lists. The total weighted student response rate was 96 percent. The institution participation rates were generally lowest among for-profit institutions and institutions whose highest offering is less than a 4-year program.

For the *student record abstraction* phase of the study (referred to as CADE for studies prior to NPSAS:12), institution completion rates were 94 percent (weighted) for institutions choosing field-CADE in NPSAS:08, approximately 96 percent for institutions choosing self-CADE, and 98 percent for data-CADE (submitting data via electronic files). CADE completion rates varied by type of institution, ranging from 92 percent for private nonprofit less-than-2-year institutions to 100 percent for private nonprofit less-than-4-year institutions. Overall, the student-level CADE completion rate (the percentage of NPSAS-eligible sample members for whom a completed CADE record was obtained) was 96 percent (weighted). Weighted student-level completion rates ranged from 87 percent for private, nonprofit, less-than-4-year institutions to 99 percent for public, 4-year, non-doctorate-granting institutions. Weighted completion rates by student type were 96 percent for undergraduate and 97 percent for graduate and first-professional students.

Overall, 95,360 of approximately 132,800 eligible sample members (72 percent unweighted) completed either a full or partial NPSAS:08 student interview. The weighted response rate was 71 percent overall and ranged from 56 percent for private, for-profit, less-than-2 year institutions to 77 percent for public, 4-year, doctorate-granting institutions.

For NPSAS:12, the unweighted institution response rate was 88 percent, while the unweighted interview completion rate was 69 percent. Across institution level and control, student response rates ranged from 55 percent for private for-profit less-than 2-year institutions to 82 percent for private nonprofit 4-year doctorate-granting institutions. Potential FTBs were significantly less likely to respond than other undergraduates (60 percent compared with 73 percent), with graduate and professional students (83 percent) completing at a higher rate than undergraduate students (66 percent).

For NPSAS:16, the unweighted institution response rate was 93 percent, while the unweighted interview completion rate was 66 percent. Across institution level and control, student response rates ranged from 48 percent for private for-profit less-than 2-year institutions to 73 percent for private nonprofit 4-year doctorate-granting institutions. Potential B&B sample members were significantly more likely to respond than other undergraduates (67 percent compared with 63 percent), with graduate students (73 percent) completing at a higher rate than undergraduate students (64 percent).

Item nonresponse. Each NPSAS institution is unique in the type of data it maintains for its students. Because not all desired information is available at every institution, the CADE software allows entry of a "data not available" code. In NPSAS:08, item response rates student record abstraction were very high overall. Two items had low response rates: marital status (46 percent) and additional phone numbers (17 percent). Thus, student records frequently lack these items. The other items had response rates ranging from 73 percent to just below 100 percent.

Missing data for items in the NPSAS:08 student interview were associated with several factors: (1) a true refusal to answer, (2) an unknown answer, (3) confusion over the question wording or response options, or (4) hesitation to provide a "best guess" response. Item nonresponse rates were based on the number of interview respondents to whom the item was applicable and of whom it was asked. Overall, item-level nonresponse rates were low, with only 23 items out of approximately 500 having more than 10 percent of data missing.

For NPSAS:12, the item-level nonresponse analysis showed that of 364 interview items, 11 items had more than 10 percent missing data.

For NPSAS:16, the item-level nonresponse analysis showed that of the approximately 500 interview items, four items had more than 10 percent missing data.

Measurement error. Due to the complex design of NPSAS, there are several possible sources of measurement error, as described below.

Sources of response. Each source of information in NPSAS has both advantages and disadvantages. While students are more likely than institutions to have a comprehensive picture of education financing, they may not remember or have records of exact amounts and sources. This information may be more accurate in student financial aid records and government databases since it is recorded at the time of application for aid.

Institutional records. While financial aid offices maintain accurate records of certain types of financial aid provided at their own institution, these records are not necessarily inclusive of all support and assistance. They may not maintain records of financial aid provided at other institutions attended by the student, and they may not include employee educational benefits and institutional assistantships, which are often treated as employee salaries. These amounts are assumed to be underreported.

Government databases. Federal aid information can only be extracted from federal financial aid databases if the institution can provide a valid Social Security number for the student. It is likely that there is some undercoverage of federal aid data in NPSAS.

CATI question delivery and data entry. Any deviation from item wording that changes the intent of the question or obscures the question meaning can result in misinterpretation on the part of the interviewee and an inaccurate response. CATI entry error occurs when the response to a question is recorded incorrectly. Measures of question delivery and data entry are used for quality assurance monitoring. Due to ongoing monitoring of student telephone interviews, problems are usually detected early and the CATI interviewers are retrained, if necessary. Overall error rates in NPSAS:08 were low (typically below 2 percent) and within control limits.

Self-administered web survey. Self-administration introduces challenges not experienced with single-mode interviewer-administered surveys. For instance, in self-administration, interviewers are not able to clarify question intent and probe when responses are unclear. Surveys also 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. These considerations were addressed in the design of the survey, making the two modes as consistent as possible.

Data Comparability

As noted above, important design changes have been implemented in NPSAS across administrations. While sufficient comparability in survey design and instrument was maintained to ensure that comparisons with past NPSAS studies could be made, institution eligibility conditions have changed since the inception of the NPSAS studies in three notable ways. First, beginning with NPSAS:2000, an institution had to be eligible to distribute federal Title IV aid to be included. Next, institutions that offered only correspondence courses—provided that these same institutions were also eligible to distribute federal Title IV student aid—were first included in NPSAS:04. Finally, institutions in Puerto Rico were not originally included in NPSAS in 1987, but subsequently were added to administrations of NPSAS between 1993 and 2008 and in 2016. Institutions in Puerto Rico are not included in 2012 administration of NPSAS. Puerto Rican institutions enroll only about 1 percent each of undergraduate and graduate students nationally. These institutions have unique aid, enrollment, and demographic patterns that distinguish them from institutions in the 50 states and the District of Columbia. Analysts wishing to compare NPSAS:12 to other NPSAS administrations may filter those data sets to exclude Puerto Rico. NCES resumed the inclusion of Puerto Rican institutions in NPSAS:16.

Comparisons with IPEDS Data. Revised weights for NPSAS:08 were released simultaneously with the release of NPSAS:12 data. NCES has reweighted NPSAS:08 data to match weighting procedures used in NPSAS:12. At the time NPSAS:08 was originally released, only 2006–07 12-month enrollment counts were available from the Integrated Postsecondary Education Data System (IPEDS) for poststratification (weighting estimates to known population totals). The revised weights, which use 2007–08 12-month enrollment counts, provide better estimates in sectors where significant enrollment shifts occurred between 2006–07 and 2007–08. Prior NPSAS iterations did not use IPEDS 12-month enrollment counts for poststratification and, as such, are unaffected.

NCES recommends that readers not try to produce their own estimates (e.g., the percentage of all students receiving aid or the numbers of undergraduates enrolled in the fall who receive federal or state aid) by combining estimates from NPSAS publications with IPEDS enrollment data. The IPEDS enrollment data are for fall enrollment only and include some students not eligible for NPSAS (e.g., those enrolled in U.S. Service Academies

and those taking college courses while enrolled in high school).

Table NPSAS-1. Weighted response rates for NPSAS administrations: Selected years, 1996 through 2016.

Component	Institution list participation rate	Student response rate	Overall
NPSAS:96			
Student survey (analysis file[1])	91	96	88
Student survey (student interview)	91	76	70
NPSAS:2000			
Student survey (analysis file[1])	91	97	89
Student survey (student interview)	91	72	66
NPSAS:04			
Student survey (analysis file[1])	80	91	72
Student survey (student interview)	80	71	56
NPSAS:08			
Student survey (analysis file[1])	90	96	86
Student survey (student interview)	90	71	64
NPSAS:12[2]			
Student survey (analysis file[1])	87	91	79
Student survey (student interview)	87	69	60
NPSAS:16[2]			
Student survey (analysis file[1])	90	93	84
Student survey (student interview)	90	66	59

— Not available

[1]The NPSAS analysis file contains analytic variables derived from all NPSAS data sources (including institutional records and extant data sources) as well as selected direct student interview variables.

[2]Study members, not interview respondents, are the unit of analysis in NPSAS:12 and NPSAS:16.

NOTE: The student interview response rates for NPSAS:96 and NPSAS:2000 are for CATI interviews only. The response rates for student interviews in NPSAS:04 include all interview modes.

SOURCE: Methodology reports for the National Postsecondary Student Aid Study. Reports are available at https://nces.ed.gov/pubsearch/getpubcats.asp?sid=013

6. CONTACT INFORMATION

For content information on NPSAS, contact:

Tracy Hunt-White
Phone: (202) 245-6507
E-mail: Tracy.Hunt-White@ed.gov

Mailing Address
National Center for Education Statistics
Institute of Education Sciences
Potomac Center Plaza
550 12th Street, SW
Washington, DC 20202

7. METHODOLOGY AND EVALUATION REPORTS

General

Berkner, L.K., He, S., Lew, S., Cominole, M.B., and Siegel, P.H. (2005). *2003–04 National Postsecondary Student Aid Study (NPSAS:04) Student Financial Aid Estimates for 2003–04* (NCES 2005-158). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. http://nces.ed.gov/pubs2005/2005158.pdf.

Cominole, M., Riccobono, J., Siegel, P., and Caves, L. (2010). *2007–08 National Postsecondary Student Aid Study (NPSAS:08) Full-scale Methodology Report* (NCES 2011-188). U.S. Department of Education. Washington, DC: National Center for Education Statistics. http://nces.ed.gov/pubs2011/2011188a.pdf.

Cominole, M.B., Siegel, P.H., Dudley, K., Roe, D., and Gilligan, T. (2006). *2004 National Postsecondary Student Aid Study (NPSAS:04) Full-Scale Methodology Report* (NCES 2006-180). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. http://nces.ed.gov/pubs2006/2006180.pdf.

Flyer, P.A. (1987). Finite Population Correction for Replication Estimates of Variance. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*. Alexandria, VA: American Statistical Association.

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*.

Kott, P.S. (1988). Model-Based Finite Population Correction for the Horvitz-Thompson Estimator. *Biometrika, 75*(4): 797–799.

Loft, J.D., Riccobono, J.A., Whitmore, R.W., Fitzgerald, R.A., and Berkner, L.K. (1995). *Methodology Report for the National Postsecondary Student Aid Study, 1992–93* (NCES 95-211). National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. http://nces.ed.gov/pubs95/95211.pdf.

Malizio, A.G. (1990). *Methodology Report for the National Postsecondary Student Aid Study, 1987* (NCES 90-309). National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=90309.

Malizio, A.G. (1992). *Methodology Report for the 1990 National Postsecondary Student Aid Study* (NCES 92-080). National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. http://nces.ed.gov/pubs92/92080.pdf.

Malizio, A.G. (1995). *Electronic E.D. TAB: National Postsecondary Student Aid Study: Estimates of Student Financial Aid: 1992-93* (NCES 95-746). National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. https://nces.ed.gov/pubs95/95746.pdf.

Malizio, A.G. (1997). *National Postsecondary Student Aid Study, 1995-96: Student Financial Aid Estimates for Federal Aid Recipients, 1995-96* (NCES 97-937). National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. https://nces.ed.gov/pubs/97937.pdf.

Malizio, A.G. (2001). *National Postsecondary Student Aid Study: Student Financial Aid Estimates for 1999–2000* (NCES 2001-209). National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. http://nces.ed.gov/pubs2001/2001209.pdf.

Radwin, D., Conzelmann, J.G., Nunnery, A., Lacy, T.A., Wu, J., Lew, S., Wine, J., and Siegel, P. (2018). *2015–16 National Postsecondary Student Aid Study (NPSAS:16): Student Financial Aid Estimates for 2015–16* (NCES 2018-466). U.S. Department of Education. Washington, DC: National Center for Education Statistics. https://nces.ed.gov/pubs2018/2018466.pdf.

Radwin, D., Wine, J., Siegel, P., and Bryan, M. (2013). *2011–12 National Postsecondary Student Aid Study (NPSAS:12): Student Financial Aid Estimates for 2011–12* (NCES 2013-165). U.S. Department of Education. Washington, DC: National Center for Education Statistics. https://nces.ed.gov/pubs2013/2013165.pdf.

Riccobono, J.A., Cominole, M.B., Siegel, P.H., Gabel, T.J., Link, M.W., and Berkner, L.K. (2002). *National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000) Methodology Report* (NCES 2002-152). National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. http://nces.ed.gov/pubs2002/2002152.pdf.

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). National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. http://nces.ed.gov/pubs98/98073.pdf.

Wei, C.C., Berkner, L., He, S., Lew, S., Cominole, M., and Siegel, P. (2009). *2007–08 National Postsecondary Student Aid Study (NPSAS:08): Student Financial Aid Estimates for 2007–08: First Look* (NCES 2009-166). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. http://nces.ed.gov/pubs2009/2009166.pdf.

Wine, J., Bryan, M., and Siegel, P. (2014). *2011–12 National Postsecondary Student Aid Study (NPSAS:12) Data File Documentation* (NCES 2014-182). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. http://nces.ed.gov/pubs2014/2014182.pdf.

Wine, J., Siegel, P., Stollberg, R. (2018) 2015–16 *National Postsecondary Student Aid Study (NPSAS:16) Data File Documentation* (NCES 2018-482). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. https://nces.ed.gov/pubs2018/2018482.pdf.

Survey Design

Malizio, A.G. (1996). *National Postsecondary Student Aid Study: 1996 Field Test Methodology Report* (NCES Working Paper 96-17). National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. http://nces.ed.gov/pubs96/9617.pdf.

Data Quality and Comparability

Ault, K., Black, S., Chromy, J., Fahimi, M., Siegel, P.H., Trofimovich, L., Whitmore, R.W., and Berkner, L.K. (2003). *Imputation Methodology for the National Postsecondary Student Aid Study: 2004* (NCES Working Paper 2003-20). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. http://nces.ed.gov/pubs2003/200320.pdf.

Salvucci, S., Walter, E., Conley, V., Fink, S., and Saba, M. (1997). *Measurement Error Studies at the National Center for Education Statistics* (NCES 97-464). National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. http://nces.ed.gov/pubs97/97464.pdf.