2011-12 NATIONAL POSTSECONDARY STUDENT AID STUDY (NPSAS:12)

STUDENT INTERVIEW AND STUDENT RECORDS

Supporting Statement Parts B and C
Request for OMB Review
(OMB # 1850-0666 v.10)

Submitted by
National Center for Education Statistics
U.S. Department of Education

October 14, 2011

B.   Collection of Information Employing Statistical Methods

1.   Respondent Universe

The students eligible for inclusion in the NPSAS:12 full scale sample are those who were enrolled in a NPSAS-eligible institution in any term or course of instruction at any time between July 1, 2011 and April 30, 2012 and who were:

- enrolled in *either* (a) an academic program; (b) at least one course for credit that could be applied toward fulfilling the requirements for an academic degree; (c) exclusively non-credit remedial coursework but who the institution has determined are eligible for Title IV aid; *or* (d) 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; and

- not currently enrolled in high school; and

- not enrolled solely in a GED or other high school completion program.

NPSAS-eligible institutions are required during the relevant academic year (2011-12 for the full-scale study) to:

- offer an educational program designed for persons who have completed secondary education; and

- offer at least one academic, occupational, or vocational program of study lasting at least 3 months or 300 clock hours; and

- offer courses that are open to more than the employees or members of the company or group (e.g., union) that administers the institution; and

- have a signed Title IV participation agreement with the U.S. Department of Education;

- be located in the 50 states or the District of Columbia; and

- be other than a U.S. Service Academy.

Institutions providing only a vocational, recreational, or remedial courses or only in-house courses for their own employees are excluded. U.S. Service Academies are excluded because of their unique funding/tuition base. The eligibility requirements are consistent with those used in all previous NPSAS rounds, with three exceptions: (1) the last requirement was new for NPSAS:2000; (2) offering more than just correspondence courses was no longer a requirement beginning with NPSAS:04; and (3) Puerto Rico has been excluded from the sample.[1]

2.   Statistical Methodology

This document presents detailed design specifications for the NPSAS:12 full-scale institutional and student samples. Section 2.a of this document describes how we created the institutional frame and selected the institutional sample. Section 2.c outlines our plans for the student sample design and selection of the student sample as well as quality control checks for both the student enrollment lists and the sampling procedures.

---

[1] After extensive review, Puerto Rico was removed from the national sample for NPSAS:12. Since the sample is national, not at the jurisdiction level, no information about PR students is lost as a result of this decision. Moreover, because postsecondary education in PR is so dissimilar to that in the 50 states, the presence of PR postsecondary students in the sample was found to skew key estimates for Hispanic student populations nationally.
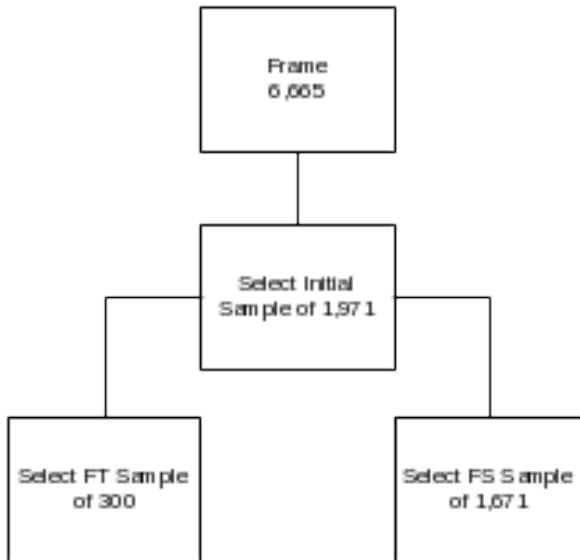
a. Institution Frame and Sample Selection

To be eligible for NPSAS:12, an institution will be required, during the 2011-12 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 are 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 or the District of Columbia;

- be other than a U.S. Service Academy; and

- have a signed Title IV participation agreement with the U.S. Department of Education.

Institutions providing only avocational, recreational, or remedial courses or only in-house courses for their own employees will be excluded. U.S. Service Academies are excluded because of their unique funding/tuition base.

The NPSAS:12 full-scale institution sampling frame was constructed during the field test from the IPEDS:2008–09 header, Institutional Characteristics (IC), 12-Month and Fall Enrollment, and Completions files. For the small number of institutions on the frame that had missing enrollment information, we imputed the data using the latest IPEDS imputation procedures to guarantee complete data for the frame.

We selected the field test institution sample statistically, rather than purposively as had been done in past NPSAS cycles. A statistical sample provides more control to ensure that the field test and the full-scale institution samples have similar characteristics, and will allow inferences to be made to the target population, supporting the analytic needs of the field test experiments. In order to accomplish this, NPSAS:12 also changed the process by which the institution sample was selected. Previous cycles selected the full-scale sample prior to selecting the field test sample from the complement. NPSAS:12 selected both institution samples simultaneously. First, a sample of 1,971 institutions, comprising the institutions needed for both the field test and full-scale studies, was selected from the stratified frame. Then, 300 of the 1,971 institutions were selected for the field test using simple random sampling within institutional strata. The remaining 1,671 institutions comprise the full-scale sample. Figure 1 displays the flow of institution sampling activities.

Figure 1. NPSAS:12 institution sample flow

```
                    ┌──────────────┐
                    │    Frame     │
                    │    6,665     │
                    └──────┬───────┘
                           │
                    ┌──────┴───────┐
                    │ Select Initial│
                    │ Sample of 1,971│
                    └──────┬───────┘
              ┌────────────┴────────────┐
      ┌───────┴────────┐        ┌───────┴────────┐
      │ Select FT Sample│       │ Select FS Sample│
      │    of 300       │       │    of 1,671     │
      └────────────────┘        └────────────────┘
```

We selected institutions for the initial sample using sequential probability minimum replacement (pmr) sampling (Chromy 1979), which resembles stratified systematic sampling with probabilities proportional to a composite measure of size (Folsom, Potter, and Williams 1987). This is the same methodology that we have used since NPSAS:96. PMR allows institutions to be selected multiple times but, instead of allowing that to happen, all institutions with a probability of being selected more than once were instead included in the sample one time with certainty, i.e., were a certainty institution. Institution measures of size were determined using annual enrollment data from the most recent IPEDS 12-Month and Fall Enrollment Components. Using composite measure of size sampling will ensure that target sample sizes are achieved within institution and student sampling strata while also achieving approximately equal student weights across institutions.

We freshened the institution sample in order to add newly eligible institutions to the sample and produce a sample that is representative of institutions eligible in the 2011-2012 academic year. To do this, we used the IPEDS:2009‑10 header, Institutional Characteristics (IC), 12-Month and Fall Enrollment, and Completions files to create an updated sampling frame of currently NPSAS-eligible institutions. This frame was then compared with the original frame, and 387 new or newly eligible institutions were identified. These 387 institutions make up the freshening sampling frame. The freshening sample size was then determined such that the freshened institutions would have similar probabilities of selection to the originally selected institutions within sector (stratum) in order to minimize unequal weights and subsequently variances. 21 freshened institutions were selected.

The 10 institutional strata are based on institutional level, control, and highest level of offering:[2]

1. public less-than-2-year
2. public 2-year
3. public 4-year non-doctorate-granting
4. public 4-year doctorate-granting
5. private nonprofit less-than-4-year
6. private nonprofit 4-year non-doctorate-granting
7. private nonprofit 4-year doctorate-granting
8. private for-profit less-than-2-year
9. private for-profit 2-year
10. private for-profit 4-year.

Although prior NPSAS administrations aggregated private for-profit 2-year and 4-year institutions into one sampling strata, the two will be split in NPSAS:12 into separate strata to reflect the recent growth in enrollment in for-profit 4-year institutions.

We expect to obtain about an overall 99 percent eligibility rate among sampled institutions. The institutional response rate is expected to be about 85 percent. The eligibility and response rates will likely vary by institutional strata. The institution sample sizes by stratum were determined such that the sampling rates would be similar to those in NPSAS:04, which was the last NPSAS to spin off a BPS Study. Additionally, we are planning to sample additional for-profit institutions. About 45 percent of the institutions on the sampling frame are for-profit, so the proportion of for-profit institutions in the sample was increased from the originally planned 17 percent to 25 percent, as shown in table 1, which presents expected rates, the institution sample sizes and estimated sample yield by the ten institutional strata.

Within each institutional stratum, additional implicit stratification for the full-scale was accomplished 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] (3) Carnegie classifications of postsecondary institutions;[4] (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); (5) 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 (6) the institution measure of size. The objective of this implicit stratification was to approximate proportional representation of institutions on these measures.

---

[2] The institutional strata can be aggregated by control or level of the institution for the purposes of reporting institution counts.
[3] The Hispanic-serving institutions (HSI) indicator no longer exists in IPEDS, so an HSI proxy was created using IPEDS Hispanic enrollment data.
[4] Some Carnegie categories were collapsed for the purposes of implicit stratification.

Table 7.   NPSAS:12 full-scale institution sample sizes and estimated yield

Institutional sector	Frame count[1]	Number sampled	Number eligible	List respondents
Total	7,052	1,692	1,672	1,416
Public				
Less-than 2-year	271	22	19	14
2-year	1,108	381	381	335
4-year non-doctorate-granting	356	130	130	117
4-year doctorate-granting	309	230	230	200
Private				
Nonprofit less-than-4-year	263	20	20	17
Nonprofit 4-year non-doctorate-granting	1,031	260	260	218
Nonprofit 4-year doctorate-granting	555	221	221	183
For-profit less-than-2-year	1,513	55	52	41
For-profit 2-year	1,028	115	111	90
For-profit 4-year	618	258	248	201

[1] Institution counts based on IPEDS:2008-09 and IPEDS:2009-10 header files.

b.   Student Sample Design and Selection

The students eligible for inclusion in the sample are those who are enrolled in a NPSAS-eligible institution in any term or course of instruction between July 1, 2011, and April 30, 2012, and who are:

- enrolled in *either* (a) an academic program; (b) at least one course for credit that could be applied toward fulfilling the requirements for an academic degree; (c) exclusively non-credit remedial coursework but who the institution **has** determined are eligible for Title IV aid; *or* (d) 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 currently enrolled in high school; and

- not enrolled *solely* in a GED or other high school completion program.

In this section, we describe the student sample design, including our plans for sampling students from enrollment lists.

Based on past experience, we expect to obtain, minimally, an overall 95 percent student eligibility rate and an overall 70 percent student interview response rate. The preliminary sample sizes and sample yield are presented in table 2. As indicated in the table, the sample will be designed to include about 125,000 students. The distribution of the sample by institution and student strata will be finalized after we complete identification of key analytic domains, especially for FTB's.

Several student subgroups will be intentionally sampled at rates different than their natural occurrence within the population due to specific analytic objectives. We anticipate that two groups will be oversampled to increase our ability to better understand their unique experiences within postsecondary education. Specifically:

    (a) Undergraduates, both FTB and non-FTB, at all award levels enrolled in for-profit institutions, who receive about 25 percent of disbursed federal aid despite constituting only about 11 percent of the student population; and

    (b) FTB undergraduates enrolled in sub-baccalaureate programs at all types of institutions, who have important early labor market experiences that can only be explored via BPS if a sufficient starting sample is identified.

Similarly, we anticipate that two student groups will be undersampled: graduate students in business and graduate students in education. Because of their sheer numbers, these sample members make it difficult to draw inference about the experiences of graduate students in other disciplines, particularly those related to science, engineering, technology, and mathematics (STEM), which we will also oversample.

We will identify potential FTBs for longitudinal follow-up, and the remaining undergraduate students will be classified as other undergraduates. The NPSAS sampling rates for students identified as potential FTBs and other undergraduate students will be adjusted based on field test interview and pre-sampling matching (see below) results, as well as on results from both NPSAS:04 and BPS:04/06, to yield the appropriate sample sizes after accounting for expected false positive and false negative rates by sector. Table 2 does not include the adjusted sample sizes, but a large percentage of the sample may be comprised of potential FTBs in order to obtain a BPS:12/14 sample yield of at least 18,300.

Table 8. NPSAS:12 preliminary student sample sizes and yield

Institutional sector	Sample students				Eligible students				Interview Respondents				Responding students per responding institution
	Total	FTBs	Other undergraduate Students	Graduate students	Total	FTBs	Other undergraduate students	Graduate students	Total	FTBs	Other undergraduate students	Graduate students	
Total	124,644	45,413	63,497	15,735	118,748	43,007	60,400	15,340	83,124	31,045	41,746	10,333	59
Public													
Less-than 2-year	1,279	717	563	0	1,040	583	458	0	630	370	260	0	46
2-year	41,304	14,822	26,483	0	38,836	13,936	24,900	0	26,534	10,127	16,408	0	79
4-year non-doctorate-granting	8,288	1,954	4,962	1,372	8,145	1,920	4,876	1,349	6,165	1,571	3,853	742	53
4-year doctorate-granting	20,057	3,864	11,182	5,012	19,801	3,814	11,039	4,948	15,388	3,187	8,791	3,410	77
Private nonprofit													
Less-than-4-year	1,646	937	709	0	1,489	847	642	0	821	492	329	0	48
4-year non-doctorate-granting	8,411	2,350	3,458	2,604	8,247	2,304	3,391	2,553	6,406	1,927	2,764	1,715	29
4-year doctorate-granting	8,770	2,100	1,733	4,937	8,533	2,080	1,699	4,754	6,524	1,722	1,426	3,376	36
Private for-profit													
Less-than-2-year	9,937	5,263	4,674	0	8,689	4,643	4,047	0	4,449	2,524	1,925	0	109
2 year	7,650	4,367	3,283	0	7,366	4,205	3,161	0	5,005	2,979	2,026	0	56
4 year	17,302	9,041	6,450	1,810	16,600	8,675	6,189	1,737	11,202	6,146	3,966	1,090	56

NOTE: FTB = first time beginner.

The eleven student sampling strata will be:

1. first-time beginning undergraduate students enrolled in sub-baccalaureate programs

2. other first-time beginning undergraduate students

3. other undergraduate students

4. master's degree students in STEM programs

5. master's degree students in education and business programs

6. master's degree students in other programs

7. doctoral-research/scholarship/other students in STEM programs

8. doctoral-research/scholarship/other students in education and business programs

9. doctoral-research/scholarship/other students in other programs

10. doctoral-professional practice students[5]

11. other graduate students[6]

We have confirmed that we have sufficient sample for the graduate student strata. We have increased the sample size for graduate students to help offset the increased design effect and variance for analyses of all graduate students.

As was done in past rounds of NPSAS, the eleven student strata will be sampled at different rates to control the sample allocation. Differential sampling rates facilitate obtaining the target sample sizes necessary to meet analytic objectives for defined domain estimates.

At the present time, we plan to employ a variable-based (rather than source-based) definition of a study member, like that used in NPSAS:04 and NPSAS:08, updated as needed by any changes to the interview. Specifically, a NPSAS:12 study member will be defined as any sample member who is determined to be eligible for the study, has a completed student interview and/or student record abstraction, and, minimally, has valid data from any source for the following:

- student type (undergraduate or graduate);

- 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;

---

[5] Past rounds of NPSAS have included samples of first-professional students. However, IPEDS has replaced the term first-professional with doctoral-professional practice. We will work with the sample institutions when requesting enrollment lists to ensure that they understand how to identify doctoral-research/scholarship/other and doctoral-professional practice students.
[6] "Other graduate" students are those who are not enrolled in a degree program, such as students just taking graduate courses.

- class level;

- first-time beginner (FTB) status;

- months enrolled;

- tuition;

- received federal aid;

- received non-federal aid;

- student budget;

- race; and

- parent education.

We expect the rate of study membership to be about 90 percent.[7]

Creating student sampling frames. Sample institutions will be asked to provide an electronic student enrollment list. The following data items will be requested for NPSAS-eligible students enrolled at each sample institution. Most of these items are the same as what was collected in past NPSAS studies:

- Name

- Social Security Number (SSN)

- Student ID number (if different from SSN)

- Student level (undergraduate, masters, doctoral-research/scholarship/other, doctoral-professional practice, other graduate)

- First-time beginner (FTB) indicator

- Class level of undergraduates (first year, second year, etc.)

- Date of birth (DOB)

- CIP or major

- Degree program

- High school graduation date (month and year)

- Contact information (local and permanent street address and phone number and school and home e-mail address)

As with NPSAS:04 and NPSAS:08, we will request locating data from institutions concurrent with the collection of student lists used for sample selection. This will allow web-based student record collection and interviewing to begin almost immediately after sample selection and thus help meet the tight schedule for data collection, data processing, and file development. For institutions unwilling to provide locating data for all students on enrollment

---

[7] NPSAS has many administrative data sources, along with the student interview. Key variables have been identified across the various data sources to determine the minimum requirements to support the analytic needs of the study. Sample members who meet these minimum requirements will be classified as *study members*. These study members will have enough information from these multiple sources to be included in the NPSAS analysis files.

lists, we will request locating data only for sampled students immediately after the sample is selected.

The FTB indicator, student level, class level, and date of birth will be used to identify and oversample potential FTBs, as described below.

High school graduation date has not been requested on lists in the past for NPSAS, so we tested the feasibility of this request in the field test. About 80 percent of institutions were able to provide this field, and no institution complained about being asked to do so. The information proved useful in identifying current high school students who were ineligible for the study.

CIP code and major have been collected in the past for NPSAS in order to help identify baccalaureate recipients who are business majors, so that they could be undersampled in NPSAS years that spin off the Baccalaureate and Beyond Longitudinal Study. In NPSAS:08, CIP code and major were also used to oversample STEM majors who were not SMART grant recipients. For NPSAS:12, as described above, we will undersample business and education graduate students and oversample graduate students in STEM fields. Therefore CIP code and major will be collected in the full-scale study. Schools will also be asked to provide degree program which we will use to identify FTBs in sub-baccalaureate programs, as described above.

In the field test, we requested an indicator of whether the institution received an ISIR (electronic record summarizing the results of the student's FAFSA processing) from CPS. This was considered potentially useful in FTB analyses; however, it has not proved useful for that purpose and, therefore, will not be requested in the full-scale study.

Obtaining student enrollment lists. To ensure the secure transmission of sensitive information on the enrollment lists, we will provide the following options to institutions: (1) upload encrypted student enrollment list files to the project's secure website using a login ID and "strong" password provided by RTI, or (2) provide an appropriately encrypted list file via e-mail (RTI will provide guidelines on encryption and creating "strong" passwords). In the field test, only two institutions e-mailed their lists, and the rest of the institutions uploaded them.

Based on NPSAS:08 and field test results we expect that few institutions will ask to provide a paper list. However, in the event that an institution is unable to transmit data via the secure electronic methods of transmission outlined above, we will accept faxes sent to a secure electronic fax machine. To ensure the fax transmission is sent to the appropriate destination, we will require a test run with nonsensitive data prior to submission of the transcripts to eliminate errors in transmission from misdialing. RTI will provide institutions with a fax cover page that includes a confidentiality statement to use when transmitting individually identifiable information.

List files received via e-fax are stored as electronic files on the e-fax server, which is housed in a secured data center at RTI. These files will be copied to a project folder that is only accessible to project staff members. Access to the project folder will be set so that only those who have authorized access will be able to see the included files. After being copied, the files will be deleted from the e-fax server. The files will be stored on the network that is backed up regularly to avoid the need to recontact the institution to provide the data again should a loss occur. RTI's information technology service (ITS) will use standard procedures for backing up data, so the backup files will exist for 3 months.

Identifying FTBs during the base year. Accurately qualifying sample members as FTBs is important because unacceptably high rates of misclassification (i.e.., false positives) can and have resulted in (1) excessive cohort loss with too few eligible sample members to sustain the longitudinal study, (2) excessive cost to "replenish" the sample with little value added, and (3) inefficient sample design (excessive oversampling of "potential" FTBs) to compensate for anticipated misclassification error.

We will take several steps early in the NPSAS:12 listing and sampling processes to improve the rate at which FTBs are correctly classified for sampling. First, in addition to an FTB indicator, we will request that enrollment lists provided by institutions (or institution systems) include class level, student level, date of birth, and high school graduation date. Students identified by the institution as FTBs, but also identified as in their fourth year or higher and/or not an undergraduate student, will not be classified as FTBs for sampling. Additionally, students appearing to be dually-enrolled at the postsecondary institution and in high school based on the high school graduation date will also not be eligible for sampling. If the FTB indicator is not provided for a student on the list but the student is 18 years old or younger and does not appear to be dually-enrolled, the student will be classified as an FTB for sampling. Otherwise, if the FTB indicator is not provided for a student on the list and the student is over the age of 18, then the student will be sampled as an "other undergraduate," but will be part of the BPS cohort if identified during the interview as an FTB.

Second, prior to sampling, we will match all students listed as potential FTBs to National Student Loan Data System (NSLDS) records to determine if any have a federal financial aid history pre-dating the NPSAS year (earlier than July 1, 2011). Since NSLDS maintains current records of all Title IV grant and loan funding, any students with data showing disbursements from prior years can be reliably excluded from the sampling frame of FTBs. Given that about 60 percent of FTBs receive some form of Title IV aid in their first year, this matching process will not be able to exclude all listed FTBs with prior enrollment, but will significantly improve the accuracy of the listing prior to sampling, yielding fewer false positives. All potential FTBs will be sent to NSLDS because ten percent of students 18 and younger sampled as FTBs and interviewed in the field test were not FTBs (false positives). In the field test, matching to NSLDS identified about 19 percent of the cases sent for matching as false positives. The field test showed that it is feasible to send all potential FTBs to NSLDS for matching. NSLDS has a free process to match the FTBs, and lists were usually returned to us in one day.

Third, simultaneously with NSLDS matching, we will match all potential FTBs to the Central Processing System (CPS) to identify students who, on their FAFSA, indicated that they had attended college previously. In the field test, we evaluated this process for potential FTBs from a subset of 94 institutions, mainly public and private nonprofit institutions, and found that we identified as false positives an additional 2.4 percent of the initial pool of potential FTBs who were not identified by NSLDS and NSC. CPS has an automated, free process for matching that we have used for other purposes in the past for NPSAS sample students. This matching can handle large numbers of cases, and the matching usually takes one day. Because there is a cost for matching to another source, described below, we plan to continue matching to CPS.

Fourth, after NSLDS and CPS matching, we will match a subset of the remaining potential FTBs to the National Student Clearinghouse (NSC) for further narrowing of FTBs based on the presence of evidence of earlier enrollment. In the field test, matching to NSC identified about 14 percent of the remaining potential FTBs, after NSLDS matching, as false

positives. NSC worked with us to set up a process that can handle a large number of potential FTBs and return FTB lists to us within two or three days. There is a "charge per case matched" for NSC matching, so we plan a targeted approach to the matching. We plan to target potential FTBs over the age of 18 in the public 2-year and for-profit sectors because these sectors had high false-positive rates in the field test and NPSAS:04 and have large full-scale sample sizes. Additional targeting and subsetting may be needed depending on budget.

Fifth, in setting our FTB selection rates, we will take into account error rates observed in the field test, NPSAS:04, and BPS:04/06 and the expected pre-sampling matching results, within each sector, as described above. As shown in table 3, some NPSAS:04 institution sectors were better able to accurately identify their students as FTBs. While the sample selection rates will take into account false positive rates, we do anticipate achieving an improvement in accuracy from the NSLDS, CPS, and NSC record matches. Table 4 shows the field test false positive identification from NSLDS, NSC, CPS, and overall matching, as well as from the interview, by sector.

Table 9. Weighted false positive rate observed in FTB identification, by sector: NPSAS:04

Sector in NPSAS:04	False positive rate (weighted)
Public	
Less-than 2-year	64.4
2-year	72.5
4-year non-doctorate-granting	26.8
4-year doctorate-granting	27.0
Private nonprofit	
Less-than-4-year	63.1
4-year non-doctorate-granting	43.4
4-year doctorate-granting	15.2
Private for-profit	
Less-than-2-year	63.1
2 years or more	70.0

FTB = first time beginner.

Table 10. Unweighted false positive rate observed in FTB identification from NSLDS, NSC, CPS, and overall matching and from the interview, by sector: NPSAS:12 field test

Sector in NPSAS:12	False positive rate (unweighted)				
	NSLDS	NSC	CPS	Overall matching	Interview
All institutions	19.2	14.1	10.2	32.1	18.2
Public					
Less-than 2-year	29.8	29.1	0	50.2	53.8
2-year	22.0	13.7	14.9	34.9	18.1
4-year non-doctorate-granting	6.7	7.0	9.5	15.6	8.4
4-year doctorate-granting	4.2	13.0	6.4	18.2	6.4
Private nonprofit					
Less-than-4-year	8.7	16.3	14.0	26.3	24.1
4-year non-doctorate-granting	20.1	12.3	6.0	31.2	10.4
4-year doctorate-granting	13.0	14.1	8.4	26.3	7.9
Private for-profit					
Less-than-2-year	33.6	25.4	0	50.4	18.8
2 year	25.5	20.8	0	41.0	44.4
4-year	48.4	22.7	9.1	60.2	28.6

FTB = first time beginner.

Quality control checks for lists. Several checks on the quality and completeness of student lists will be implemented before the sample students are selected. For example, the lists will fail quality control checks if student level and/or the FTB indicator are not included on the list. Additionally, the unduplicated total of students at the undergraduate and graduate levels on each institution's student list will be checked against the latest IPEDS unduplicated enrollment data from the 12-Month Enrollment Component. The unduplicated count of FTBs on each institution's student list will be checked against IPEDS enrollment data from the Fall Enrollment Component adjusted to estimate full year enrollment. Contact information will be checked carefully for each enrollment list as well as for each student sampled. If an institution does not provide high school graduation date, but includes many students who are less than 18, the list will fail quality control checks because it may erroneously contain high school students.

Institutions failing quality control checks will be re-contacted to resolve the discrepancy and verify that the institution coordinator who prepared the student list(s) clearly understood our request and provided a list of the appropriate students. Should we determine that the initial list provided by an institution is not satisfactory, we will request a replacement list. We will proceed with selecting sample students when we have either confirmed that the list received is correct or have received a corrected list. If the list is incorrect, but the institution will not or cannot correct

it, we will determine if we can proceed with selecting sample students, depending on the problem.

Selection of sample students. Students will be sampled on a flow basis, as student lists are received, using a stratified systematic sampling procedure. Sample yield will be monitored by institution and student sampling strata, and the sampling rates will be adjusted early, if necessary, to achieve the desired sample yield.

Quality control checks for sampling. RTI has developed technical operating procedures (TOPs) that describe how to properly implement statistical procedures and QC checks. We will employ a checklist for use by all NPSAS:12 statisticians to ensure that appropriate QC checks are performed.

Some specific sampling QC checks will include, but will not be limited to, checking that the:

- students on the sampling frames all have a known, non-zero probability of selection;
- number of students selected match the target sample size; and
- sample weight for each student is the inverse of the probability of selection.

Tracing prior to the start of data collection. Once the sample is selected, RTI will conduct several batch database searches to prepare the sample for the start of student interviews. The first steps in the batch tracing process will be to match to the U.S. Department of Education's Central Processing System (CPS). After CPS matching, FirstData's National Change of Address (NCOA) and Phone Append service will be used to obtain updated contact information. Any new information collected from CPS, NCOA, or Phone Append matches will be added to the NPSAS locator database Batch tracing is the final step before the start of data collection.

3. Methods for Maximizing Response Rates

Response rates in the NPSAS:12 full-scale study are a function of success in two basic activities: identifying and locating the sample members involved, then contacting them and gaining their cooperation. Two classes of respondents are involved: institutions and students who were enrolled in those institutions. Institutions will be asked to provide data from institutional records for sampled students. In this section, we describe our plans for maximizing response to the request for data from institutional records. We also present our plans for maximizing response to the student survey.

a. Collection of Data from Institutional Records

Our plans for contacting and communicating with institutions, beginning with the process of list acquisition, are designed to ensure the cooperation of as many institutions as possible and to establish rapport with institutional staff. This process will include sending the chief administrator of each institution a package of descriptive materials about the study, follow-up telephone calls to obtain the chief administrator's consent and cooperation, and asking the chief administrator to designate an Institutional Coordinator (IC) who will serve as our primary point of contact.

All institution coordinators receive information that informs them about the purposes of NPSAS, describes their tasks, and assures them of our commitment to maintaining the confidentiality of data. Written materials will be provided to coordinators explaining each phase of the study, as well as their role in each. These contacts and activities, as well as enrollment list collection, were included and approved in the NPSAS:12 institution contacting and enrollment list collection OMB package.

Training of institution coordinators is geared toward the method of data collection selected by the institution (see below). The system used for collecting institutional record data is a World Wide Web application; and the website, accessible only with an ID and password, provides institution coordinators with instructions for all phases of study participation. Copies of all written materials, as well as answers to frequently asked questions, are available on the website.

Experienced RTI staff from RTI's Call Center Services (CCS) carry out these contacts and are assigned to specific institutions, which remain their responsibility throughout the data collection process. This allows RTI staff members to establish rapport with the institution's staff and provides those individuals with a consistent point of contact at RTI. Staff members are thoroughly trained in basic financial aid concepts and in the purposes and requirements of the study, which helps them establish credibility with the institution staff. As an additional means of maximizing institutional participation, we have secured endorsements from 26 professional associations for NPSAS:12 (see Appendix G).

RTI will offer several options for providing the Student Records for sampled students (as in prior NPSAS studies), and invite the coordinator to select the methodology that is least burdensome and most convenient for the institution. The optional methods for providing student record data are:

Student Records obtained via a web-based data entry interface. The web-based data entry interface is flexible and allows the coordinator to enter data in one of two data entry modes. One data entry mode resembles a spreadsheet (referred to as "grid mode") and as such, the coordinator can view and edit multiple student records at a time. The other data entry mode displays one student at a time, and the coordinator may enter data in a top to bottom fashion before moving onto the next student.

Student Records obtained by completing an Excel workbook. An Excel workbook will be created for each institution and will be preloaded with the sampled students' ID, name, and SSN (if available). To facilitate simultaneous data entry by different offices within the institution, the workbook contains a separate worksheet for each of the following topic areas: Financial Aid, Enrollment, Locating and Contact info, Demographics. The user will download the Excel worksheet from the secure NPSAS institution website, enter the data, and then upload the data to the website. Validation checks occur both within Excel as data are entered and when the datas are uploaded via the website.

Student Records obtained by uploading CSV (comma separated values) files. Institutions with the means to export data from their internal database systems to a flat file may opt for this method of supplying Student Records. Over the past two NPSAS studies, the number of institutions providing data files has increased. Institutions that select

this method will be provided with detailed import specifications, and all data uploading will occur through the project's secure website.

Institution coordinators will receive a guide that provides detailed instructions for accessing and using the website. A video tutorial covering how to provide data via the website will also be available on the website. The guide and the script for the tutorial are provided in Appendix J.

Prior to data collection, student records are matched to the U.S. Department of Education Central Processing System (CPS)—which contains data on federal financial aid applications—for locating purposes and to reduce the burden on the institutions for the student record abstractions. The vast majority of the federal aid applicants (about 95 percent) will match successfully to the CPS prior to Student Records data collection. During data collection, institutions will be asked to provide the student's last name and Social Security number for the small number of federal aid applicants who did not match to the CPS on the first attempt. After Student Records data collection ends, we will submit the new names and Social Security numbers to CPS for file matching. Any new data obtained for the additional students will be delivered on the Electronic Code Book (ECB) with the data obtained prior to Student Records data collection.

b. Student Survey: Online and Telephone Interviews (CATI)

Methods for maximizing response to the study survey include: (1) tracing of sample members; (2) thorough training for all staff involved in data collection; (3) use of a sophisticated case management system; (4) a carefully designed survey instrument; and (5) detailed plans for averting and converting refusals.

1. Tracing of Sample Members

To achieve the desired response rate, we propose an integrated tracing approach that consists of up to 11 steps designed to yield the maximum number of locates with the least expense. The steps of our tracing plan include the following elements.

- *Matching student list information with CPS, NCOA, Phone Append, and other databases,* which will yield the most current locating information for the students sampled for NPSAS:12.

- *Providing a system for moving locating information obtained during collection of student record data quickly into the locator database* so that this new information can be put to immediate use in CATI.

- *Lead letter and other mailings as necessary to sample members.* A personalized letter (signed by an NCES official) and study brochure will be mailed to all sample members to initiate data collection. This letter will include a toll-free number, study website address, and Study ID and password, and will request that sample members complete the self-administered interview over the Internet. A subset of students least likely to participate in the NPSAS:12 interview will move directly to the Call Center to begin outbound calling immediately. A few days after the lead letter mailing, an email message mirroring the letter will also be sent to sample members. Additional mailings and emails will occur throughout data collection to prompt sample members to participate.

- *Conducting batch tracing before data collection and after the start of CATI as needed.* Not all schools will be able to give complete or up-to-date locating information on each student, and some cases will require more advanced tracing, before mailings can be sent or the cases can be worked in CATI. RTI plans to conduct batch tracing on all cases to obtain updated address information prior to mailing the lead letters. This step will minimize the number of returned letters and maximize the number of early completes. To handle cases for which no mailing address, phone number, or other contact information is available, RTI plans to conduct advance tracing of the cases prior to lead letter mailout and data collection. This advance tracing will involve searching for address and telephone information. As lead information is found, additional searches will be conducted through interactive databases to expand on leads found. This will be an important step in the tracing components because of the nature of this sample. After locating information is found, more advanced database searches, such as Experian, will be used, to provide more comprehensive information for the individual.

- *CATI tracing.* Telephone interviewers (TIs) will call available telephone numbers to attempt to contact the sample member. If the sample member is no longer available at a particular telephone number, TIs will probe for any additional locating information.

- *Pre-intensive tracing using Premium Phone.* To minimize the number of cases requiring more expensive intensive interactive tracing, we will send cases to Premium Phone to identify a new phone number. Through Premium Phone we can tap into 475 million landline, VOIP and wireless records, including over 380 million self reported names, addresses, and telephone numbers. Premium Phone can also tap into 70 percent of cell phones in the U.S.; obtaining reliable cell phone numbers is becoming an increasingly critical component of locating and interviewing this population.

- *Conducting intensive in-house tracing, including proprietary database searches, including postsecondary institution web sites and social networking pages.* RTI's tracing specialists conduct intensive interactive searches to locate contact information for sample members. In NPSAS:08, about 60 percent of sample members requiring intensive tracing were located, and about 59 percent of those located responded to the interview. Intensive interactive tracing differs from batch tracing in that a tracer can assess each case on an individual basis to determine which resources are most appropriate and the order in which they should be used. Intensive interactive tracing is also much more detailed due to the personal review of information. During interactive tracing, tracers utilize all previously obtained contact information to make tracing decisions about each case. These intensive interactive searches are completed using a special program that works with the locator database to provide organization and efficiency in the intensive tracing process.

- *Conducting NPSAS List Completer (NLC) searches.* NLC is an RTI software application that compiles all information available for the school and sample members to Tracing Services for additional address, phone, and e-mail searches to

be made. The application will send Tracing Services the school name, school web address, and total number of students to be worked. If student name, address, and phone number are available, this information will also be sent to the NLC. Tracing Services will then use the school web page directly to conduct searches and update records of student information that can be accessed from the web.

- *Providing an online video to introduce the study.* Using stop-motion filmmaking and Lego figures, RTI will develop a video to communicate important study information to sample members.

2.   Training for Data Collection Staff

Telephone data collection will be conducted by staff in RTI's Call Center Services unit, including Quality Control Supervisors (QCSs), Help Desk Agents (HDAs), Telephone Interviewers (TIs), and Refusal Conversion Specialists. Training programs for these staff members are critical to maximizing response rates and collecting accurate and reliable data.

Quality control supervisors, who are responsible for all supervisory tasks, will attend project-specific training for QCSs, in addition to the content of the HDA and TI training. They will receive an overview of the study, background and objectives, and the data collection instrument through a question-by-question review. Supervisors will also receive training in the following areas: providing direct supervision during data collection; handling refusals; monitoring interviews and maintaining records of monitoring results; problem resolution; case review; specific project procedures and protocols; reviewing CATI reports; and monitoring data collection progress.

Training for HDAs, who assist sample members who call the project-specific toll-free line, and Telephone Interviewers is designed to help staff become familiar with and practice using the Help Desk application and survey instrument, as well as to learn project procedures and requirements. Particular attention will be paid to quality control initiatives, including refusal avoidance and methods to ensure that quality data are collected. Both HDAs and TIs will receive project-specific training on telephone interviewing, and HDAs will receive additional training specifically geared toward solving technical problems and answering questions from web participants regarding the study or related to specific items within the interview. They will also be able to reissue passwords and respond to sample member e-mail messages, using prepared text approved by NCES. At the conclusion of training, all HDAs and TIs must meet certification requirements by successfully completing a certification interview. This evaluation consists of a full-length interview with project staff observing and evaluating interviewers, as well as an oral evaluation of interviewers' knowledge of the study's Frequently Asked Questions.

3.   Case Management System

Student interviews will be conducted using a single web-based survey instrument for both self-administered and CATI data collection. The data collection activities will be accomplished through the Case Management System (CMS), which is equipped with the following capabilities:

- on-line access to locating information and histories of locating efforts for each case;

- state-of-the-art questionnaire administration module with full "front-end cleaning" capabilities (i.e., editing as information is obtained from respondents);

- sample management module for tracking case progress and status; and

- automated scheduling module which delivers cases to interviewers and incorporates the following features:

- *Automatic delivery of appointment and call-back cases at specified times*. This reduces the need for tracking appointments and helps ensure the interviewer is punctual. The scheduler automatically calculates the delivery time of the case in reference to the appropriate time zone.

- *Sorting of non-appointment cases according to parameters and priorities set by project staff*. For instance, priorities may be set to give first preference to cases within certain sub-samples or geographic areas; cases may be sorted to establish priorities between cases of differing status. Furthermore, the historic pattern of calling outcomes may be used to set priorities (e.g., cases with more than a certain number of unsuccessful attempts during a given time of day may be passed over until the next time period). These parameters ensure that cases are delivered to interviewers in a consistent manner according to specified project priorities.

- *Restriction on allowable interviewers*. Groups of cases (or individual cases) may be designated for delivery to specific interviewers or groups of interviewers. This feature is most commonly used in filtering refusal cases, locating problems, or foreign language cases to specific interviewers with specialized skills.

- *Complete records of calls and tracking of all previous outcomes*. The scheduler tracks all outcomes for each case, labeling each with type, date, and time. These are easily accessed by the interviewer upon entering the individual case, along with interviewer notes, thereby eliminating the need for a paper record of calls of any kind.

- *Flagging of problem cases for supervisor action or supervisor review*. For example, refusal cases may be routed to supervisors for decisions about whether and when a refusal letter should be mailed, or whether a supervisor or other more experienced refusal converter should be assigned to make the next call.

- *Complete reporting capabilities*. These include default reports on the aggregate status of cases and custom report generation capabilities.

The integration of these capabilities reduces the number of discrete stages required in data collection and data preparation activities and increases capabilities for immediate error reconciliation, which results in better data quality and reduced cost. Overall, the scheduler provides a highly efficient case assignment and delivery function by reducing supervisory and clerical time, improving execution on the part of interviewers and supervisors by automatically monitoring appointments and call-backs, and reducing variation in implementing survey priorities and objectives.

4. Survey Instrument Design

In January 2010, NCES received approval to conduct focus groups – the first stage in a multistage qualitative evaluation comprising focus groups and cognitive interviews. This qualitative evaluation informed refinement of items used in previous surveys as well as the development of items which elaborate the postsecondary choices of the first-time beginning (FTB) population. Focus groups were conducted to help the instrument design team move from conceptualization to instrument development. Additionally, the focus groups were used to improve a select set of existing questions in the NPSAS interview, particularly items involving financial aid terminology that is possibly unfamiliar to students (e.g. private loans) and items used to determine eligibility for the BPS cohort. Cognitive testing was conducted prior to both the field test and full-scale data collections. Results of the focus groups and cognitive testing have guided the development of final instrument wording.

To specify and program the survey, NPSAS:12 employs a cutting-edge web-based instrument and deployment system, created by RTI, known as Hatteras. NPSAS:08 was the first NCES study to use Hatteras, a flexible, collaborative system allowing for instrument designers, programmers, and NCES management to work together on instrument development. The instrument specifications stored in database tables are used to produce web pages dynamically. Hatteras provides multimode functionality, whereby the survey instrument is created one time and can be used for self-administration, CATI, CAPI, or data entry. Hatteras provides multilanguage support and is compatible with RTI's Instrument Development and Documentation System (IDADS) to provide NCES-specific data documentation directly from the instrument specifications.

Below are some of the basic questionnaire administration features of the web-based instrument:

- Based on responses to previous questions, the respondent or interviewer is automatically routed to the next appropriate question, according to predesignated skip patterns.

- The web-based interview automatically inserts "text substitutions" or "text fills" where alternate wording is appropriate depending on the characteristics of the respondent or his/her responses to previous questions.

- The web-based interview can incorporate or preload data about the individual respondent from outside sources (e.g., previous interviews, sample frame files, administrative data, etc.). Such data are often used to drive skip patterns or define text substitutions. In some cases, the information is presented to the respondent for verification or to reconcile inconsistencies.

- With the web/CATI instrument, numerous question-specific probes may be incorporated to explore unusual responses for reconciliation with the respondent, to probe "don't know" responses as a way of reducing item non-response, or to clarify inconsistencies across questions.

- An innovative improvement to previous NPSAS data collections, the web-based instrument uses an assisted coding mechanism to code text strings provided by

respondents. Drawing from a database of potential codes, the assisted coder derives a list of options from which the interviewer or respondent can choose an appropriate code (or codes if it is a multi-level variable with general, specific, and/or detail components) corresponding to the text string.

- When identical sets of questions will be repeated for an unidentified number of entities, such as children, jobs, schools, and so on, the system allows respondents to cycle through these questions as often as is needed.

In addition to the functional capabilities of the CMS and web instrument described above, our efforts to achieve the desired response rate will select from among established procedures proven effective in other large-scale studies we have completed, including earlier NPSAS collections. These include:

- Providing multiple response modes, including self-administered and interviewer-administered options.

- Offering incentives to all sample members to encourage response (for NPSAS:12, the amount will be fixed at $30).

- Strategic use of prompting calls initiated prior to the start of data collection to remind sample members about the study and the importance of their participation.

- Assigning experienced telephone interviewers who have proven their ability to contact and obtain cooperation from a high proportion of sample members.

- Training interviewers thoroughly on study objectives, study population characteristics, and approaches that will help gain cooperation from sample members.

- Providing interviewing staff with a comprehensive set of questions and answers that will provide encouraging responses to questions that sample members may ask.

- Maintaining a high level of monitoring and direct supervision so that interviewers who are experiencing low cooperation rates are identified quickly and corrective action is taken.

- Making every reasonable effort to obtain an interview following the initial data collection announcement, but allowing respondent flexibility in scheduling appointments to be interviewed.

- Providing hesitant respondents with a toll-free number to use to telephone RTI and discuss the study with the project director or other senior project staff.

- Thoroughly reviewing all refusal cases and making special conversion efforts whenever feasible (see next section).

5. Refusal Aversion and Conversion

Recognizing and avoiding refusals is important to maximize the response rate. We will emphasize this and other topics related to obtaining cooperation during data collector training. Supervisors will monitor interviewers intensely during the early days of data collection and provide retraining as necessary. In addition, the

supervisors will review daily interviewer production reports produced by the CATI system to identify and retrain any data collectors who are producing unacceptable numbers of refusals or other problems.

After encountering a refusal, the data collector enters comments into the CMS record. These comments include all pertinent data regarding the refusal situation, including any unusual circumstances and any reasons given by the sample member for refusing. Supervisors will review these comments to determine what action to take with each refusal. No refusal or partial interview will be coded as final without supervisory review and approval. In completing the review, the supervisor will consider all available information about the case and will initiate appropriate action.

If a follow-up is clearly inappropriate (e.g., there are extenuating circumstances, such as illness or the sample member firmly requested that no further contact be made), the case will be coded as final and will not be recontacted. If the case appears to be a "soft" refusal, follow-up will be assigned to an interviewer other than the one who received the initial refusal. The case will be assigned to a member of a special refusal conversion team made up of interviewers who have proven especially adept at converting refusals.

Refusal conversion efforts will be delayed for at least one week to give the respondent some time after the initial refusal. Attempts at refusal conversion will not be made with individuals who become verbally aggressive or who threaten to take legal or other action. Refusal conversion efforts will not be conducted to a degree that would constitute harassment. We will respect a sample member's right to decide not to participate and will not impinge this right by carrying conversion efforts beyond the bounds of propriety.

6.   Other means of increasing response. RTI will draw upon its most effective strategies for increasing survey participation among the main sample as well as important subgroups. The following strategies may be used for NPSAS:12:

- Cases in sectors that have tended to not participate online during the early completion period will be sent directly for outbound telephone interviewing, with the option to complete the interview online. The additional 3 weeks available at the start of data collection will provide extra contacting time before sample members relocate during the summer months.

- RTI will send more frequent and/or special mailings and emails to target subgroups of cases that need additional prompting.

- Typically, cases that refuse or are reluctant to complete the full interview are offered an abbreviated interview. For NPSAS:12, the abbreviated interview will be offered to low propensity sectors, when appropriate, to ensure adequate representation of the sector.

- As needed, sectors will be allocated additional intensive tracing time and/or use supplementary tracing resources.

- Help Desk staff and telephone interviewers will be provided better resources to assist sample members interested in completing the interview on the web. Interviewers will be able to send immediate emails to sample members providing the website address and the ID/password needed to access the interview.   In addition,, if RTI receives permission to send text messages to sample members, they will send periodic text messages as prompts.

- If response rates are lagging significantly in low propensity sectors, NCES will discuss with OMB additional options for increasing their participation.

4.   Tests of Procedures and Methods

There will be no tests of procedures or methods as part of the full-scale NPSAS:12 student interview and student records collection. The procedures and methods were tested as part of the field trial conducted in 2011 (1850-0666 v.8).

5.   Reviewing Statisticians and Individuals Responsible for Designing and Conducting the Study

Names of individuals consulted on statistical aspects of study design along with their affiliation and telephone numbers are provided below.

Name	Affiliation	Telephone
Dr. Susan Choy	MPR	(510) 849-4942
Ms. Christina Wei	MPR	(510) 849-4942
Dr. John Riccobono	RTI	(919) 541-7006
Dr. James Chromy	RTI	(919) 541-7019
Mr. Peter Siegel	RTI	(919) 541-6348
Dr. Jennifer Wine	RTI	(919) 541-6870

In addition to these statisticians and survey design experts, the following statisticians at NCES have also reviewed and approved the statistical aspects of the study: Dr. Tom Weko, Dr. Tracy Hunt-White, Dr. Matt Soldner, Dr. Sean Simone, and Mr. Ted Socha.

6. Other Contractors' Staff Responsible for Conducting the Study

The study is being conducted by the Postsecondary, Adult, and Career Education (PACE) division of the National Center for Education Statistics (NCES), U.S. Department of Education. NCES's prime contractor is the RTI International (RTI). RTI is being assisted through subcontracted activities by MPR Associates, Branch Associates, Kforce Government Solutions, Inc. (KGS), Research Support Services, Millennium Services 2000+, Inc., and consultants. Principal professional staff of the contractors, not listed above, who are assigned to the study are identified below:

Name	Affiliation
Dr. Cynthia Decker	Consultant
Ms. Andrea Sykes	Consultant
Mr. Dan Heffron	KGS
Ms. Bart Ecker	Millennium Services
Ms. Vicky Dingler	MPR
Ms. Laura Horn	MPR
Ms. Alexandria Radford	MPR
Dr. Jennie Woo	MPR
Dr. Alisú Shoua-Glusberg	RSS
Mr. Jeff Franklin	RTI
Ms. Christine Rasmussen	RTI
Ms. Kristin Dudley	RTI
Mr. Brian Kuhr	RTI

C. Overview of Analysis Topics and Survey Items

The two NPSAS:12 data collection instruments (including a facsimile of the student interview and a table showing data elements to be collected from student records) are presented in Appendixes H and I. Many of the data elements to be used in NPSAS:12 appeared in the previously approved NPSAS:04 and NPSAS: 96 studies, the last NPSAS studies to include a BPS cohort. Additional items will also be included in NPSAS:12. As described above, these items were tested with focus groups and cognitive testing prior to both the field test and full-scale data collections.

NPSAS is a particularly complex survey because it uses a large variety of sources and several sources may be available for the same data element. These sources include:

- Student records: Student-level data from institutional records collected through a secure web application. These include records from the registrar, bursar, and financial aid office.

- Student interviews: Data from student interviews using either the web-based self-administered or telephone interview.

- ACT: Data from American College Testing service files on ACT college entrance examinations and student questionnaires.

- FAFSA: Data from the Central Processing System (CPS) for Free Application for Federal Student Aid (FAFSA), which includes student and parent demographic, income and asset information, and expected family contribution used in need analysis. The records are called Institutional Student Information Records (ISIR).

- IPEDS: Data from the Integrated Postsecondary Education Data System (IPEDS) which includes institutional characteristics and enrollment.

- NSLDS: Data from the U.S. Department of Education's National Student Loan Data system (NSLDS), which has a record of all individual student loans ever borrowed and all Pell Grant payments since 1994.

- NSC: Data from the National Student Clearinghouse's Student Tracker file, which includes student-level data on institutions attended, enrollment dates, and degree completion.

- SAT: Data from the College Board files of SAT college entrance examinations.