

Appendix E
Sample Design for the 2008 National
Postsecondary Student Aid Study (NPSAS:08)

Sampling Design for the NPSAS:08 Full-Scale Study Collection of Information Employing Statistical Methods

All procedures, methods, and systems to be used in the full-scale study were tested in a realistic operational environment during the field test conducted during the 2006–07 academic year. Specific plans for full-scale activities are provided below.

E.1 Respondent Universe

E.1.1 Institution Universe

To be eligible for the NPSAS:08 full-scale study, institutions are required during the 2007–08 academic year to:

- offer an educational program designed for persons who have 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 administers the institution;
- have a signed Title IV participation agreement with the U.S. Department of Education;
- be located in the 50 states, the District of Columbia, or Puerto Rico; and
- be other than a U.S. Service Academy.

Institutions providing only avocational, 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.

E.1.2 Student Universe

The students eligible for inclusion in the sample for the NPSAS:08 full-scale study are those who were enrolled in a NPSAS-eligible institution in any term or course of instruction at any time from July 1, 2007 through April 30, 2008 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; *or* (c) 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.

E.2 Statistical Methodology

E.2.1 Sample design and proposed augmentations

The details describing the design and allocations of the institutional and student samples are presented in sections E.2.2 and E.2.3. This first section describes two augmentations to the sample design as it was originally proposed.

The first augmentation involves oversampling 5,000 recipients of SMART grants and/or Academic Competitiveness Grants (ACG) (two new sources of student financial aid), to ensure that these students are sufficiently well represented for analysis. RTI will establish sampling rates for SMART grant recipients from a file that is to be provided by ED no later than December 2007. After establishing sampling rates, we will use the ED file to flag SMART grant recipients on lists provided by institutions.

More students are expected to receive ACG than SMART grants, so an oversample of ACG recipients may not be necessary. We will look at sample sizes with and without oversampling and at the effects of oversampling on variance estimates. In consultation with NCES we will decide if an ACG oversample is necessary. If oversampling of ACG recipients is not necessary, then the additional sample of 5,000 students will be only for SMART grant recipients.¹

The second augmentation is contingent upon 1) funding of a pending proposal to the Department Of Education and 2) a planned modification to that proposal based on discussions with the Commissioner of the National Center for Education Statistics as well as attendees of the recent Technical Review Panel meeting (held 8/28-29, 2007). The NPSAS:08 full-scale sample will be augmented to include state-representative samples of undergraduate students in four sectors from six states which will make it possible to produce state-level analyses and comparisons of many of the most pertinent issues in postsecondary financial aid and prices.²

As originally designed, the NPSAS:08 sample yields estimates that are nationally representative but generally not large enough to permit comparison of critical subsets of students within a particular state. Tuition levels for public institutions (attended by about 80 percent of all undergraduates) vary substantially by state, as does the nature of state grant programs (i.e., large versus small, need-based versus merit-based). Therefore, it is possible to analyze the effect of these policies and programs with federal and institutional financial aid policies and programs only at the state level.

The choice of states for the sample augmentation was based on several considerations, including

- **Size of undergraduate enrollments in four sectors:** public 4-year, private not-for-profit 4-year, public 2-year, and private for-profit, degree-granting institutions. We estimate that we will need approximately 1,200 respondents per

¹ The sample design described below assumes that SMART grant recipients will be oversampled and ACG recipients will not be oversampled.

² The field test institutional sample was selected from the complement of institutions selected for the full-scale study to avoid asking an institution to participate in both. After field test data collection, ED requested that RTI augment the full-scale sample to provide state-level representation of students in selected states and sectors. To accomplish this goal, it will be necessary to include a small number of institutions that participated in the field test in the full-scale study.

state in the 4-year and for-profit sectors and 2,000 respondents in the public 2-year sector in order to yield a sufficient number of full-time, dependent, low-income undergraduates—the subset of students that is of particular relevance for the study of postsecondary access. Tuition and grant policies in the states with the largest enrollments have the greatest effect on national patterns and trends. As a practical matter, their representation in a national sample is already so large that the cost of sample augmentation is relatively low.

- **Prior inclusion in the NPSAS:04 12-state sample and high levels of cooperation and participation in that survey.** Participation in NPSAS is not mandatory for institutions, so we depend on institutional cooperation within a state to achieve the response rates and yields required for reliable estimates. Smaller states that were willing and helpful in NPSAS:04 and achieved high yields and response rates are more likely to cooperate again, and with less effort.
- **States with different or recent changes in tuition and state grant policies that provide opportunities for comparative research and analysis.**

Using these criteria, we proposed to augment the samples for the following 6 states: California, Texas, New York, Illinois, Georgia, and Minnesota.

The sample sizes presented in this appendix reflect the inclusion of the SMART grant oversample and the state-representative samples. The institution sampling strata will be expanded to include strata for the four sectors within each of the six states. For selecting institutions within states and sectors, there are three scenarios. First, for some sectors in the states, there are already enough institutions in the sample, so that no additional sample institutions are necessary. In this case, the institutions already selected will stay in sample. Second, for other sectors in the states, all institutions in the sector in the state will be in sample. Therefore, the institutions already selected will remain in the sample, and the remaining institutions will be added to the sample. Third, for other sectors in the state, additional institutions need to be added to the sample, but not all institutions will be selected. In this case, the originally selected institutions are no longer necessarily in sample, and a new sample will be selected. This is the cleanest method statistically and is also best to keep the unequal weighting effect (UWE) from being too large. In the second and third scenarios, it is anticipated that a total of about 20 field test sample institutions may be included in the full-scale sample.

Also, the student strata will be expanded to include SMART grant recipients and to include in-state and out-of-state students.

E.2.2 Institution Sample

The institution samples for the field test and full-scale studies were selected simultaneously, prior to the field test study. The institutional sampling frame for the NPSAS:08 field test was constructed from the 2004-05 Integrated Postsecondary Education Data System (IPEDS) institutional characteristics, header, completions, and fall enrollment files. Three hundred institutions were selected for the field test from the complement of institutions selected for the full-scale study to minimize the possibility that an institution would be burdened with participation in both the field test and full-scale samples, while maintaining the representativeness of the full-scale sample. However, since the decision to augment the full-scale sample to provide state-level representation of students in selected states and sectors was

made after field test data collection was completed, it will be necessary to include in the full scale study about 20 institutions that also participated in the field test (as described above).

The full-scale sample was then freshened in order to add newly eligible institutions to the sample and produce a sample that is representative of institutions eligible in the 2007-08 academic year. To do this, we used the IPEDS:2005-06 header, Institutional Characteristics (IC), 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 167 new or newly eligible institutions were identified. These 167 institutions make up the freshening sampling frame. Freshening sample sizes were 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.

Institutions were selected for the NPSAS:08 full-scale study using stratified random sampling with probabilities proportional to a composite measure of size,³ which is the same methodology that we used for NPSAS:96, NPSAS:2000, and NPSAS:04. Institution measures of size were determined using annual enrollment data from the 2004-05 IPEDS Fall Enrollment Survey and bachelor's degree data from the 2004-05 IPEDS Completions Survey. Using composite measure of size sampling ensures that target sample sizes are achieved within institution and student sampling strata while also achieving approximately equal student weights across institutions.

We expect to obtain an overall eligibility rate of 98 percent and an overall institutional participation (response) rate of 84 percent⁴ (based on the NPSAS:04 full-scale study). Eligibility and response rates are expected to vary by institutional strata. Based on these expected rates, the institution sample sizes (after freshening)⁵ and estimated sample yield, by the nine sectors traditionally used for analyses, are presented in table E-1.

³ Folsom, R.E., Potter, F.J., and Williams, S.R. (1987). Notes on a Composite Size Measure for Self-Weighting Samples in Multiple Domains. *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, 792-796.

⁴ The institution response rate of 84 percent assumes that institutional participation will not be mandatory.

⁵ The institution sampling frame was constructed from the IPEDS:2004-05 header, Institutional Characteristics, Fall Enrollment, and Completions files. 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 2007-08 academic year, using the corresponding IPEDS files for 2005-06.

Table E-1. NPSAS:08 expected full-scale estimated institution sample sizes and yield

Institutional sector	Frame count ¹	Number sampled	Number eligible	List respondents
Total	6,777	1,962	1,940	1,621
Public less-than-2-year	247	22	19	14
Public 2-year	1,167	449	449	383
Public 4-year non-doctoral	358	199	199	169
Public 4-year, doctoral	290	290	290	250
Private not-for-profit less-than-4-year	326	20	20	18
Private not-for-profit 4-year, non-doctoral	1,017	359	346	284
Private not-for-profit 4-year doctoral	591	269	269	209
Private for-profit less-than-2-year	1,476	97	91	77
Private for-profit 2-year or more	1,305	257	257	217

¹ Institution counts based on IPEDS:2004-05 header file.

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

The nine sectors traditionally used for NPSAS analyses were the basis for forming the institutional strata. These are

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 not-for-profit less-than-4-year
6. private not-for-profit 4-year non-doctorate-granting
7. private not-for-profit 4-year doctorate-granting
8. private for-profit less-than-2-year
9. private for-profit 2-year or more.

Since the NPSAS:08 student sample will be designed to include a new sample cohort for a Baccalaureate and Beyond Longitudinal Study (B&B), these nine sectors will be further broken down to form the same 22 strata used in NPSAS:2000 (the last NPSAS to generate a B&B study) in order to ensure sufficient numbers of sample students within 4-year institutions by various degree types (especially education degrees, an important analysis domain for the B&B longitudinal study). Additionally, 24 strata are necessary for the state sample, as described above. The 46 institutional sampling strata are as follows:

1. public less-than-2-year;
2. public 2-year;
3. public 4-year non-doctorate-granting bachelor's high education;
4. public 4-year non-doctorate-granting bachelor's low education;
5. public 4-year non-doctorate-granting master's high education;

6. public 4-year non-doctorate-granting master's low education;
7. public 4-year doctorate-granting high education;
8. public 4-year doctorate-granting low education;
9. public 4-year first-professional-granting high education;
10. public 4-year first-professional-granting low education;
11. private not-for-profit less-than-2-year;
12. private not-for-profit 2-year;
13. private not-for-profit 4-year non-doctorate-granting bachelor's high education;
14. private not-for-profit 4-year non-doctorate-granting bachelor's low education;
15. private not-for-profit 4-year non-doctorate-granting master's high education;
16. private not-for-profit 4-year non-doctorate-granting master's low education;
17. private not-for-profit 4-year doctorate-granting high education;
18. private not-for-profit 4-year doctorate-granting low education;
19. private not-for-profit 4-year first-professional-granting high education;
20. private not-for-profit 4-year first-professional-granting low education;
21. private for-profit less-than-2-year;
22. private for-profit 2-year or more;
23. California public 2-year;
24. California public 4-year;
25. California private not-for-profit 4-year;
26. California private for-profit degree-granting;
27. Texas public 2-year;
28. Texas public 4-year;
29. Texas private not-for-profit 4-year;
30. Texas private for-profit degree-granting;
31. New York public 2-year;
32. New York public 4-year;
33. New York private not-for-profit 4-year;
34. New York private for-profit degree-granting;
35. Illinois public 2-year;
36. Illinois public 4-year;
37. Illinois private not-for-profit 4-year;
38. Illinois private for-profit degree-granting;
39. Georgia public 2-year;
40. Georgia public 4-year;
41. Georgia private not-for-profit 4-year;
42. Georgia private for-profit degree-granting;
43. Minnesota public 2-year;

- 44. Minnesota public 4-year;
- 45. Minnesota private not-for-profit 4-year; and
- 46. Minnesota private for-profit degree-granting.

Note that “high education” refers to the 20 percent of institutions with the highest proportions of their baccalaureate degrees awarded in education (based on the most recent IPEDS Completions file). The remaining 80 percent of institutions are classified as “low education” (i.e., having a lower proportion of baccalaureate degrees awarded in education).

E.2.3 Student Sample

Based on the expected response and eligibility rates, the preliminary expected student sample sizes and sample yield are presented in table E-2. This table shows that the full-scale study will be designed to sample a total of 138,066 students, including 29,428 baccalaureate recipients; 86,274 other undergraduate students; and 22,364 graduate and first-professional students. Based on past experience, we expect to obtain, minimally, an overall eligibility rate of 92.0 percent and an overall student interview response rate of 70.0 percent; however, these rates will vary by sector.

Table E-2. NPSAS:08 preliminary full-scale student sample sizes and yield

Institutional sector	Sample students				Eligible students				Study respondents				Responding students per responding institution
	Total	Baccalaureates	Other undergraduate students	Graduate/first-professional students	Total	Baccalaureates	Other undergraduate students	Graduate/first-professional students	Total	Baccalaureates	Other undergraduate students	Graduate/first-professional students	
Total	138,066	29,428	86,274	22,364	127,073	27,827	78,026	21,220	113,178	25,567	68,110	19,501	70
Public less-than-2-year	3,409	0	3,409	0	2,719	0	2,719	0	2,238	0	2,238	0	155
Public 2-year	31,095	0	31,095	0	27,330	0	27,330	0	21,719	0	21,719	0	57
Public 4-year non-doctoral	16,592	5,722	8,710	2,153	15,739	5,430	8,266	2,043	14,139	4,878	7,425	1,835	83
Public 4-year doctoral	37,456	12,164	14,683	10,579	35,595	11,569	13,965	10,062	32,602	10,596	12,791	9,216	130
Private not-for-profit less-than-4-year	3,077	0	3,077	0	2,739	0	2,739	0	2,524	0	2,524	0	142
Private not-for-profit 4-year non-doctoral	12,577	4,752	6,065	1,734	11,783	4,461	5,694	1,628	11,091	4,199	5,360	1,532	39
Private not-for-profit 4-year doctoral	15,784	4,080	4,236	7,486	15,005	3,874	4,022	7,108	13,860	3,579	3,715	6,566	66
Private for-profit less-than-2-year	7,391	0	7,391	0	6,295	0	6,295	0	5,839	0	5,839	0	76
Private for-profit 2-year or more	10,679	2710	7,608	412	9,868	2,492	6,997	379	9,164	2,314	6,497	352	42

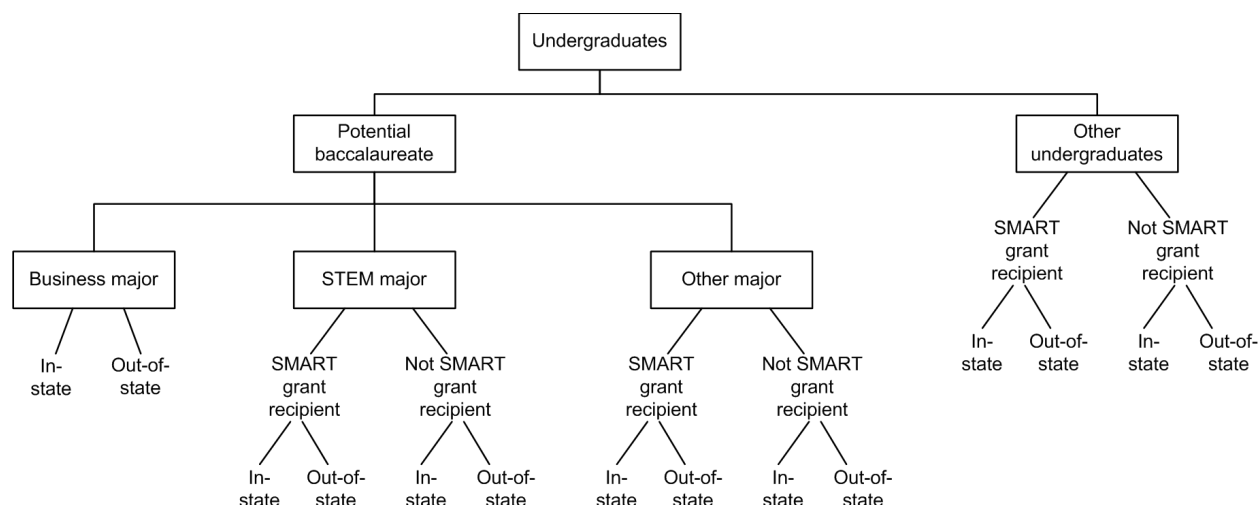
NOTE: NPSAS:08 = 2008 National Postsecondary Student Aid Study.

We plan to employ a variable-based (rather than source-based) definition of study respondent, similar to that used in the NPSAS:08 field test and in NPSAS:04. There are multiple sources of data obtained as part of the NPSAS study, and study respondents must meet minimum data requirements, regardless of source. Using the same variable-based definition from the field test, we expect the overall study response rate to be 89.1 percent, based on NPSAS:04 results. We anticipate, however, that study response rates will vary by institutional sector, as was the case in NPSAS:04. Using the rates we experienced in that study, we expect approximately 113,178 study respondents, including 25,567 baccalaureate recipients; 68,110 other undergraduate students; and 19,501 graduate and first-professional students.

The 18 student sampling strata are listed below and shown graphically in figure E-1:

1. in-state potential baccalaureate recipients who are business majors;
2. out-of state potential baccalaureate recipients who are business majors;
3. in-state potential baccalaureate recipients who are science, technology, engineering, or mathematics (STEM) majors and SMART grant recipients;
4. out-of-state potential baccalaureate recipients who are STEM majors and SMART grant recipients;
5. in-state potential baccalaureate recipients who are STEM majors and not SMART grant recipients;
6. out-of-state potential baccalaureate recipients who are STEM majors and not SMART grant recipients;
7. in-state potential baccalaureate recipients in all other majors who are SMART grant recipients;
8. out-of state potential baccalaureate recipients in all other majors who are SMART grant recipients;
9. in-state potential baccalaureate recipients in all other majors who are not SMART grant recipients;
10. out-of state potential baccalaureate recipients in all other majors who are not SMART grant recipients;
11. in-state other undergraduate students who are SMART grant recipients;
12. out-of-state other undergraduate students who are SMART grant recipients;
13. in-state other undergraduate students who are not SMART grant recipients;
14. out-of-state other undergraduate students who are not SMART grant recipients;
15. masters students;
16. doctoral students;
17. other graduate students; and
18. first-professional students.

Figure E-1. NPSAS:08 undergraduate student sampling strata



As was done in NPSAS:2000 and NPSAS:04, certain student types (potential baccalaureate recipients, other undergraduates, masters students, doctoral students, other graduate students, and first-professional students) 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 in the full-scale study.

To ensure a large enough sample for the B&B follow-up, the base year sample includes a large percentage of potential baccalaureate recipients (see table E-2). The sampling rates for students identified as potential baccalaureates and other undergraduate students on enrollment lists will be adjusted to yield the appropriate sample sizes after accounting for the baccalaureate “false-positives.” This will ensure sufficient numbers of actual baccalaureate recipients. The expected “false positive” rate will be based on the results of the NPSAS:08 field test, comparing B&B status across several sources, and on NPSAS:2000 full scale survey data.⁶

RTI will receive a file of SMART grant recipients from ED and will match that list to each institution’s enrollment list to identify and stratify such students. SMART grant recipients are required to major in a STEM field or in certain foreign languages, so baccalaureate recipients who are STEM or other majors must also be stratified by SMART grant recipient status. However, the strata for baccalaureate recipients who are business majors does not need to be stratified by SMART grant recipient status.

Creating Student Sampling Frames. Several alternatives for the types of student enrollment lists that can be provided by the sample institutions are available. Our first preference is to obtain an unduplicated list of all students enrolled in the specified time frame. However, lists by term of enrollment and/or by type of student (e.g., baccalaureate recipient, undergraduate, graduate, and first-professional) will be accepted. The student ID numbers can be used to easily unduplicate electronic files. If an institution has difficulty meeting these requirements, we will be flexible and select the student sample from whatever type of list(s) that the institution can provide, so long as it appears to accurately reflect enrollment during the specified terms of instruction. If necessary, we are even prepared to provide institutions with specifications to allow them to select their own sample.

⁶ In NPSAS:2000, the “false-positive” rate was 13 percent, but lists were usually sent closer to the end of the spring term than they will be in NPSAS:08, so this rate may be a low estimate for NPSAS:08.

In prior NPSAS studies that spun off a B&B cohort, lists of potential baccalaureate recipients were collected with the student list of all enrolled undergraduates and graduates/first professionals. Unfortunately, these baccalaureate lists often could not be provided until late in the spring or in the summer, after baccalaureate recipients could be positively identified. To help facilitate earlier receipt of lists, we will request that the enrollment lists for 4-year institutions include an indicator of class level for undergraduates (1st year, 2nd year, 3rd year, 4th year, or 5th year). From NPSAS:2000, we estimate that about 55 percent of the 4th- and 5th-year students will be baccalaureate recipients during the NPSAS year, and about 7 percent of 3rd-year students will also be baccalaureate recipients. To increase the likelihood of correctly identifying baccalaureate recipients, we will also request that the enrollment lists for 4-year institutions include an indicator (B&B flag) of students who have received or are expected to receive a baccalaureate degree during the NPSAS year (yes, no, don't know). We will instruct institutions to make this identification before spring graduation so as not to hold up the lists because of this requirement. These two indicators will be used instead of requesting a baccalaureate recipient list, and we plan to oversample 4th and 5th year undergraduates (seniors) and students with a B&B flag of "yes" to ensure obtaining sufficient yield of baccalaureate recipients for the B&B longitudinal study. We expect that most institutions will be able to provide undergraduate year for their students and a B&B flag.

We will also request major field of study and Classification of Instructional Programs (CIP) code on the lists to allow us to undersample business majors and to oversample STEM majors. A similar procedure was used effectively in NPSAS:2000 (the last NPSAS to include a B&B cohort). We expect that most institutions can and will provide the CIP codes. Undersampling business majors is necessary because a disproportionately large proportion of baccalaureate recipients are business majors, and oversampling STEM majors is necessary because there is an emerging longitudinal analytic interest in baccalaureate recipients in these fields.

The following additional data items will be requested for all NPSAS-eligible students enrolled at each sample institution:

- name;
- Date of birth (DOB);
- Social Security number (SSN);
- student ID number (if different from SSN);
- student level (undergraduate, masters, doctoral, other graduate, first-professional); and
- locating information (local and permanent street address and phone number and school and home e-mail address).

Permanent address will be used to identify and oversample undergraduate in-state students. A similar procedure was used effectively in NPSAS:04. Oversampling of in-state students in the six states with representative samples is necessary because state-level analyses typically only include in-state students, so sufficient sample size is needed. In the other states, the undergraduate students will be stratified by in-state and out-of-state for operational efficiency, but in-state students will not be oversampled.

As part of initial sampling activities, we will ask participating institutions to provide SSN and DOB for all students on their enrollment list.⁷ We recognize the sensitivity of the requested information, and appreciate the argument that it should be obtained only for sample members. However, collecting this information for all enrolled students is critical to the success of the study for several reasons:

- It is possible that some minors will be included in the study population, so we will need to collect DOB to identify minors and obtain parental consent prior to data collection.
- The NPSAS:08 study includes a special analytic focus on a new federal grant (the National SMART grant) and SSN is needed to identify and oversample recipients of this new grant.
- Having SSN will ensure the accuracy of the sample, because it is used as the unique student identification number by most institutions. We need to ensure that we get the right data records when collecting data from institutions for sampled students. It will also be used to unduplicate the sample for students who attend multiple institutions.
- Making one initial data request of institutions will minimize the burden required for participation (rather than obtaining one set of information for all enrolled students, and then later obtaining a set of information for sampled students).
- An issue related to institutional burden is institutional participation. It is very likely that some institutions will respond to the first request, but not to the second. Refusal to provide SSNs after the sample members are selected will contribute dramatically to student-level nonresponse, because it will increase the rate of unlocatable students (see the following bullet).
- Obtaining SSN early will allow us to initiate locating and file matching procedures early enough to ensure that data collection can be completed within the allotted schedule. The data collection schedule would be significantly and negatively impacted if locating activities could not begin at the earliest stages of institutional contact.
- NPSAS data are critical for informing policy and legislation, and are needed by Congress in a timely fashion. Thus, the data collection schedule is also critical. We must be able to identify the sample, locate students, and finish data collection and data processing quickly. This will not be possible within the allotted time frame if we are unable to initiate locating activities for sampled students once the sample has been selected.

The following section describes our planned procedures to securely obtain, store, and discard sensitive information collected for sampling purposes.

Obtaining student enrollment lists. The student sample will be selected from the lists provided by the sampled institutions. To ensure the secure transmission of sensitive information, we will provide the following options to institutions: (1) upload

⁷ For institutions unwilling to provide SSN or location data for all students on enrollment lists, we will request SSN or locating data only for sample students immediately after the sample is selected.

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 past administrations of this study, hard copy lists were accepted via Fed-Ex or fax. We did not offer this option in the field test and will not offer it in the full-scale study. We expect that a very few institutions will ask to provide a hard copy list (in NPSAS:04 full-scale study, 30 institutions submitted a hard-copy list—mostly via FedEx). In such cases, we will encourage one of the secure electronic methods of transmission. If that is not possible, we will accept a faxed list (but not a Fed-Ex list.) Although fax equipment and software does facilitate rapid transmission of information, this same equipment and software opens up the possibility that information could be misdirected or intercepted by individuals to whom access is not intended or authorized. To safeguard against this, as much as is practical, RTI protocol will only allow for lists to be faxed to a fax machine housed in a locked room and only if schools cannot use one of the other options. To ensure the fax transmission is sent to the appropriate destination, we will require a test run with nonsensitive data prior to submitting the actual list to eliminate errors in transmission from misdialing. RTI will provide schools with a FAX cover page that includes a confidentiality statement to use when transmitting individually identifiable information.⁸ After a sample is selected from an institution, the original electronic or keyed list of all students containing SSNs will be deleted, and faxed lists will be shredded. RTI will ensure that the SSNs for nonselected students are securely discarded (see description below).

Storage of enrollment files.

- Encrypted electronic files sent via e-mail to a secure e-mail folder will be accessible only to a few staff members on the sampling team. These files will then be copied to a project folder that is accessible only to these same staff members. Access to this project folder will be set so that only those who have authorized access will be able to see the included files. The folder will not even be visible to those without access. After being copied, the files will be deleted from the e-mail folder. After selecting the sample of students for each school, the original file containing all students with SSNs will be immediately deleted. While in use, files will be stored on the network that is backed up regularly to avoid the need to recontact the institution to provide the list 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 three months.
- Files uploaded to the secure NPSAS website will be copied from the NCES server to the same project folder mentioned above. After being moved, the files will be immediately deleted from the NCES server. After selecting the sample of students for each school, the original file containing all students with SSNs will be immediately deleted. As above, it is necessary for the files to be stored on the project share so that they can be backed up by ITS in case any problems occur that cause us to lose data. ITS will use their

⁸ These procedures are consistent with those endorsed by HIPAA. See <http://www.hipaadvisory.com/action/faxfacts.htm>

standard procedures for backing up data, so the backup files will exist for 3 months.

- Paper lists will be kept in one locked file cabinet. Only NPSAS sampling staff will have access to the file cabinet. The paper lists will be shredded immediately after the sample is selected, keyed, and QC'ed. The keying will be done by the same sampling staff who select the sample.

Selection of Sample Students. The unduplicated number of enrollees on each institution's enrollment list will be checked against the latest IPEDS unduplicated enrollment data, which are part of the spring web-based IPEDS data collection. For electronic files, lists will be unduplicated by student ID number. For faxed lists, which are expected to be small, the total number of students listed will be counted. The comparisons will be made for baccalaureates and for each student level: undergraduate, graduate, and first-professional. Based on past experience only counts within 25 percent of nonimputed IPEDs counts will pass edit. There will be one exception based on field test results: if the baccalaureate count is higher than the IPEDs count but within 50 percent, the count will pass edit because we are comparing potential baccalaureate list counts with actual IPEDs counts.

Institutions that fail edit will be recontacted to resolve the discrepancy and to verify that the institution coordinator who prepared the student lists clearly understood our request and provided a list of the appropriate students. When we determine that the initial list provided by the institution was not satisfactory, we will request a replacement list. We will proceed with selecting sample students when we either have confirmed that the list received is correct or have received a corrected list.

Electronic lists will be unduplicated by student ID number prior to sample selection. In addition, all samples, both those selected from electronic files and from paper lists, will be unduplicated by SSN *between* institutions. The duplicate sample member will be deleted from the second institution because the sample is selected on a flow basis. In prior NPSAS studies, we found several instances in which this check avoided multiple selections of the same student. However, we also learned that the ID numbers assigned to noncitizens may not be unique across institutions; thus when duplicate IDs are detected but the IDs are not standard SSNs (do not satisfy the appropriate range check), we will check the student names to verify that they are indeed duplicates before deleting the students.

Student names and SSNs or student IDs will be keyed into Excel for faxed lists, which are expected to be short lists. The keying will be checked thoroughly. These keyed lists will then be unduplicated and sampled similarly to electronic lists. After the sample is selected from a keyed list, the additional information from the original faxed list will be keyed just for the sampled students and checked carefully.

Stratified systematic samples of students will be selected, from both electronic and faxed student lists,⁹ on a flow basis as the lists are received by adapting the procedures we have used successfully for student sampling in prior NPSAS rounds. As the student samples are selected they will be added to a master sample file containing, minimally, for each sample student: a unique study ID number (NPSASID), SSN, the institution's IPEDS ID number (UNITID), institutional stratum, student stratum, and selection probability.¹⁰ Sample yield will be monitored

⁹ Based on NPSAS:04 and the field test, we expect that 2 percent or less of the enrollment lists received will be paper list.

¹⁰ The selection probability is based on the unduplicated list.

by institutional and student sampling strata, and the sampling rates will be adjusted early, if necessary, to achieve the desired full-scale sample yields.

Quality Control Checks for Sampling. All statistical procedures will undergo thorough quality control checks. We have technical operating procedures (TOPs) in place for sampling and general programming. These TOPs describe how to properly implement statistical procedures and QC checks. We will use a checklist for all statisticians to use to make sure that all appropriate QC checks are done.

Some specific sampling QC checks will include, but are not limited to, checking that

- the students on the sampling frames all have a known, non-zero probability of selection; and
- the number of students selected match the target sample sizes.