

Appendix G
NPSAS Sampling Details

Appendix G

NPSAS Sampling Details

I. Target Population

The target population for NPSAS:2000 consists of all students enrolled in Title IV participating postsecondary institutions other than U.S. Service Academies in the United States or Puerto Rico at any time during the 1999-2000 federal financial aid award year, excluding students who were enrolled solely in a GED program or who were concurrently enrolled in high school. With one exception, the survey population also was defined as those students who were enrolled at any time between July 1, 1999 and June 30, 2000. Specifically, if a term or course began after May 31, 2000 and ended after June 30, 2000, then students enrolled only in that term or course were excluded from the survey population. The target population is the population about which inferences will be made. The survey population is the population actually covered by the sampling frame. Nearly all members of the target population also are members of the survey population; however, the adopted definition of the survey population allowed the student lists needed for sample selection to be obtained before or during June for many institutions (e.g., those on a semester calendar system). More specific definitions of the institution and student populations are provided later in this appendix.

This definition of the survey population differs from previous NPSAS rounds but is more consistent with the definition of the target population. Prior NPSAS rounds also surveyed students enrolled at institutions not participating in Title IV aid programs. In addition, for NPSAS:96 and NPSAS:93, the survey population was defined as those students who were enrolled in any term beginning between May 1 and April 30 during the survey year, i.e., 1995-96 and 1992-93, respectively; for NPSAS:90, the students sampled were those enrolled on August 1, 1989, October 15, 1989, February 15, 1990, or June 15, 1990 (however, the June 15 enrollees were not sampled for 4-year institutions because of budgetary limitations); for NPSAS:87, only Fall 1986 enrollees were sampled.

II. Sample Design Overview

A schematic overview of the sequential statistical sampling process for NPSAS:2000 is provided in figure G-1. The goal of all sampling activities was to attain NCES-required numbers of eligible sample postsecondary students (within specified student and institution types). An important domain of the required student sample was the set of students identified as baccalaureates, that is, students who were enrolled and received their bachelors degree between July 1, 1999 and June 30, 2000. These students comprise the baseline cohort for the Baccalaureate and Beyond (B&B) longitudinal study. Accounting for expected (from prior NPSAS rounds) rates of nonresponse and ineligibility among sample students and rates of B&B misclassification, the desired numbers of sample students were initially determined as shown, by type of institution and type of student classification, in table G-1. Since it was necessary to select the student samples on a flow basis as sample institutions provided their enrollment lists (in order to meet the data collection schedule), the students were sampled at fixed rates. Under this approach, the actual numbers of students sampled are random variables; however, the sampling rates were set to meet or exceed, in expectation, the sample sizes shown in table G-1.

Table G-1.—Target numbers of sample students, by institutional stratum and type of student

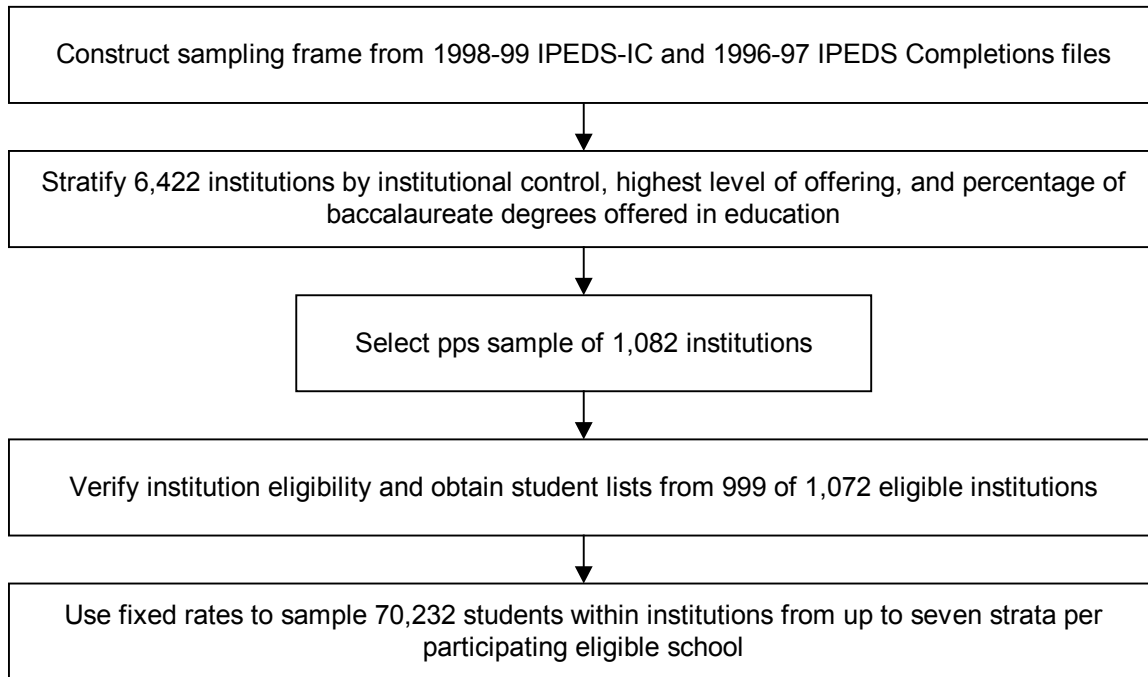
Institutional stratum	Total	Baccalaureate	Other undergraduate	Graduate	First-professional
Total	70,266	16,372	40,918	11,657	1,319
Public					
1 Less-than-2-year	1,996	†	1,996	†	†
2 2-year	10,976	†	10,976	†	†
Total less-than-4-year	12,972	†	12,972	†	†
3 Bachelor's, high education ¹	236	127	109	†	†
4 Bachelor's low education ²	923	175	740	†	†
5 Masters, high education	2,124	1,223	694	208	†
6 Masters, low education	6,640	1,970	3,636	1,042	†
Total 4-year non-doctorate-granting	9,924	3,495	5,180	1,249	†
7 Doctorate-granting, high education	2,371	1,229	719	423	†
8 Doctorate-granting, low education	5,884	1,496	2,702	1,686	†
9 First-professional-granting, high education	3,985	1,983	1,175	764	63
10 First-professional-granting, low education	9,900	2,677	4,021	2,776	427
Total 4-year doctorate-granting	22,141	7,386	8,617	5,648	490
Private, not-for-profit					
11 Less-than-2-year	601	†	601	†	†
12 2-year	1,201	†	1,201	†	†
Total less-than-4-year	1,802	†	1,802	†	†
13 Bachelors, high education	739	423	315	†	†
14 Bachelors, low education	1,586	583	999	†	†
15 Masters, high education	1,595	855	543	197	†
16 Masters, low education	3,655	1,049	1,800	810	†
Total 4-year, non-doctorate-granting	7,574	2,910	3,658	1,006	†
17 Doctorate-granting, high education	781	263	209	309	†
18 Doctorate-granting, low education	1,310	262	418	630	†
19 First-professional-granting, high education	3,216	959	1,054	994	210
20 First-professional-granting, low education	4,013	956	856	1,589	612
Total 4-year, doctorate-granting	9,320	2,439	2,538	3,521	822
Private, for-profit					
21 Less-than-2-year	4,328	†	4,328	†	†
22 2-year or more	2,203	141	1,823	232	7
Total private, for-profit	6,531	141	6,151	232	7

† Not applicable

¹A school is classified as "high-ed" if it is in the top 20 percent of its stratum in terms of baccalaureate students graduating with education degrees.²A school is classified as "low-ed" if it is not in the top 20 percent of its stratum in terms of baccalaureate students graduating with education degrees.

NOTE: First-professional-granting institutions include doctoral degrees.

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

Figure G-1.—Schematic of sequential NPSAS:2000 sampling operations

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

The NPSAS:2000 sample also was designed to achieve at least 30 student CATI respondents from each sample institution that had at least that many eligible students enrolled during the NPSAS year. This was to allow NCES to send each participating institution a report using the results of the interviews with their students without violating confidentiality requirements. Consequently, institution sample sizes were determined to achieve an average of approximately 40 or more sample students per institution within each institutional stratum. Given these student sample size goals, the desired number of participating institutions was determined to be 1,008. Based on projected institutional participation rates obtained in prior NPSAS rounds and the NPSAS:2000 field test, an initial sample of 1,082 institutions was selected.

III. The Institutional Sample

The target population for NPSAS:2000 includes nearly all Title IV participating postsecondary institutions in the 50 states, the District of Columbia, and Puerto Rico. Specifically, to be eligible for NPSAS:2000 an institution is required, during the 1999-2000 academic year, to:

- offer an educational program designed for persons who have completed secondary education
- offer more than just correspondence courses

- 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
- be located in the 50 states, the District of Columbia, or Puerto Rico
- be other than a U.S. Service Academy (which are not eligible for this financial aid study because of their unique funding/tuition base)
- 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 are excluded. The listed eligibility requirements are consistent with those used in previous NPSAS rounds, except for the last one which is new for NPSAS:2000.

A. Sample Frame Construction

The institution-level sampling frame for NPSAS:2000 was constructed from the 1998-99 Integrated Postsecondary Education Data System Institutional Characteristics (IPEDS-IC) file and the 1996-97 IPEDS Completions file. The IPEDS-IC database provides nearly complete coverage of the institutions in the target population. Listings include: (a) all institutions whose primary purpose is the provision of postsecondary education; (b) all branches of colleges, universities, and other institutions, as long as the branch offers a full program of study (not just courses); (c) free-standing medical schools, as well as schools of nursing, schools of radiology, etc., within hospitals; and (d) schools offering occupational and vocational training with the intent of preparing students for work (e.g., a modeling school training for professional modeling—not just a charm school). The IPEDS files do not include: (a) schools not open to the general public (i.e., training sites at prisons, military installations, corporations); (b) hospitals offering internships or residency programs only; or hospitals that only offer training as part of a medical school program at an institution of higher education; (c) organizational entities providing only noncredit continuing education (CEUs); (d) schools whose only purpose is to prepare students to take a particular test, (e.g., CPA examination or Bar exams); or (e) branch campuses of U.S. institutions in foreign countries. The completions file was used to obtain counts of total and business baccalaureate degree awarded and baccalaureate degrees awarded in education which, in turn, were used to compute measures of size and to stratify, respectively.

The IPEDS-IC file exclusions, themselves, eliminate some categories of ineligible institutions; however, additional deletion from this file was required. Starting with the 9,744 “institutions” on this database, records were deleted to yield a sampling frame containing 6,422 institutions *appearing* to be eligible for NPSAS:2000 based on their 1998-99 IPEDS-IC data. Deletions included: (1) administrative units; (2) U.S. Service academies; (3) schools outside of U.S. and Puerto Rico; (4) institutions offering no programs of at least 300 content hours, six semesters/trimesters, or 12 quarter hours and for which the highest level of offering was a certificate or diploma of less than one academic year; (5) Institutions offering only

correspondence courses; (6) institutions not eligible for Title IV funding; and (7) institutions selected for sample for the field test. The latter deletion was possible without compromising population coverage because the field test sample was selected using stratified simple random sampling.¹

Because enrollment data were needed to compute measures of size for sample selection, the 1998-99 IPEDS “unduplicated count” enrollment data were edited and/or imputed to eliminate missing data. Missing undergraduate, graduate, and first-professional enrollments were set to zero for institutions that did not offer that level of instruction, and missing baccalaureate counts were set to zero for institutions that did not award bachelors degrees. For institutions that provided only undergraduate instruction, missing undergraduate enrollment was obtained from the fall enrollment variables, if those were nonmissing. For institutions with any missing enrollments, enrollment was obtained from the 1997-98 IPEDS-IC file, if available. Finally, sets of records were identified for which the enrollment data either: (a) were reported with another institution’s, or (b) contained combined data. In such cases, the combined enrollment data were allocated equally to all institutions in the set. For the remaining 57 records with missing enrollment data, imputation classes (defined by institutional sector (level and control) and first-professional, graduate, and undergraduate offering (yes or no)), were created and missing enrollment data were imputed for such cases as the imputation class median. This approach avoids imputing unusually large or unusually small enrollments.

The institutions on the sampling frame were then partitioned into 22 institutional strata based on institutional control, highest level of offering, and percentage of baccalaureate degrees awarded in education:

- (1) Public less-than-2-year
- (2) Public 2-year
- (3) Public Bachelors high education²
- (4) Public Bachelors low education
- (5) Public Masters, high education
- (6) Public Masters low education
- (7) Public Doctorate-granting high education
- (8) Public Doctorate-granting low education

¹ After large institutions, which were likely to be certainty institutions in the full-scale survey, were deleted from the field test institutional sampling frame, a stratified simple random sample of institutions was selected for the field test, using the same 22 strata that were used for the full-scale study. Although no probability-based inferences were planned for the field test, a probability-based sample was used because the complement of the field test sample was used for the full-scale study sampling frame. Each institution on the full-scale sampling frame received a first-stage sampling weight based on the probability that it was not selected for the field test sample. See U.S. Department of Education. National Center for Education Statistics. National Postsecondary Student Aid Study: 2000 (NPSAS:2000) Field Test Methodology Report, NCES 2000-17, by Melissa R. Biber, Michael W. Link, John A. Riccobono and Peter H. Siegel. Washington, DC: October 2000.

² For each high education/low education breakout, the 20 percent of institutions with the highest proportions of their baccalaureate degrees awarded in education (based on the 1996-97 IPEDS Completions file) were defined to be the high education stratum. The purpose of this stratification was to ensure a certain sample size of students going into the teaching profession, which is an important analysis domain for the baccalaureate and beyond study.

- (9) Public First-professional-granting, high education
- (10) Public 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 Bachelors, high education
- (14) Private not-for-profit Bachelors, low education
- (15) Private not-for-profit Masters, high education
- (16) Private not-for-profit Masters, low education
- (17) Private not-for-profit Doctorate-granting, high education
- (18) Private not-for-profit doctorate-granting, low education
- (19) Private not-for-profit first-professional-granting, high education
- (20) Private not-for-profit first-professional-granting, low education
- (21) Private for-profit less-than-2-year
- (22) Private for-profit 2-year or more.

A stratified sample of 1,082 institutions was then selected with probabilities proportional to size (pps); some of these institutions subsequently proved to be ineligible and others failed to participate.

B. Selecting Sample Institutions

It was necessary to allocate the student sample to the separate applicable institutional (defined above) and student sampling strata. The student sampling strata used were:

- (1) students receiving a baccalaureate degree in business
- (2) students receiving a baccalaureate degree not in business
- (3) other undergraduate students
- (4) masters students
- (5) doctoral students
- (6) other graduate students
- (7) first-professional students

In determining the allocation, the following notation is used:

- (1) $r = 1, 2, \dots, 22$ indexes the previously defined institutional strata
- (2) $s = 1, 2, 3, 7$ indexes the previously defined initial student strata
- (3) $j = 1, 2, \dots, J(r)$ indexes the institutions within stratum “ r ”
- (4) $M_{rs}(j)$ = number of students enrolled during the NPSAS year who belong to student stratum “ s ” at the j -th institution in institutional stratum “ r ”
- (5) m_{rs} = number of students to be selected from student stratum “ s ” within the r -th institutional stratum (referred to henceforth as student stratum “ rs ”)
- (6) $\pi_r(j)$ = probability of selecting the j -th institution in institutional stratum “ r ”

The overall population sampling rate (f_{rs}) for student stratum “rs” is given by

$$f_{rs} = m_{rs} / M_{rs}(+)$$

where

$$M_{rs}(+) = \sum_{j=1}^{J(r)} M_{rs}(j).$$

The initially computed stratum-level student sampling rates, f_{rs} (used to define institution measures of size) are shown in table G-2.. Table G-2 presents the sampling rates for the seven student domains consisting of baccalaureate business, baccalaureate non-business, other undergraduate, masters, doctoral, other graduate, and first-professional students based on the 1998-99 IPEDS IC file and 1996-97 IPEDS Completions file counts and the required sample sizes previously presented in table G-1. The IPEDS files do not provide separate counts for masters, doctoral, and other graduate students; hence, the partitioning of total graduate enrollment into these three categories was based on NPSAS:96 data.

The composite measure of size for the j-th institution in stratum “r” was then defined to be

$$S_r(j) = \sum_{s=1}^7 f_{rs} M_{rs}(j),$$

which is the number of students that would be selected from the j-th institution if all institutions on the frame were to be sampled.

An independent sample of institutions was selected for each institutional stratum using Chromy’s³ sequential, probability minimum replacement (pmr) sampling algorithm to select institutions with probabilities proportional to their measures of size. However, rather than allow multiple selections of sample institutions, those with expected frequencies of selection greater than unity (1.00) were selected with certainty. The remainder of the institutional sample was selected from the remaining institutions within each stratum. Therefore, the probability of selection for the j-th institution in institutional stratum “r” is given by

$$\pi_r(j) = \begin{cases} \frac{n_r^* S_r(j)}{S_r(+)} & \text{for non-certainty selections,} \\ 1 & \text{for certainty selections,} \end{cases}$$

where

$$S_r(+) = \sum_{j=1}^{J(r)} S_r(j),$$

³J.R. Chromy, “Sequential Sample Selection Methods.” *Proceedings of the American Statistical Association Section on Survey Research Methods*, 1979, 401-406.

Table G-2.—Student sampling rates used in determining measures of size by institutional stratum and type of student

Institutional stratum	Undergraduate Students										
	Baccalaureate					Other undergraduates					
	Business majors		Non-business majors			Business majors		Non-business majors			
Size of universe	Sample size	Sampling rate	Size of universe	Sample size	Sampling rate	Size of universe	Sample size	Sampling rate	Size of universe	Sample size	Sampling rate
Total	221,964	1,366	0.0062	948,809	15,006	0.0158	17,121,076	40,915	0.0024		
Public less-than-2-year	†	†	†	†	†	†	†	1,996	0.0094		
Public 2-year	†	†	†	†	†	†	9,112,766	10,976	0.0012		
Total less than 4-year	†	†	†	†	†	†	9,326,108	12,972	0.0014		
Public bachelor's high education	954	14	0.0147	3,066	113	0.0369	37,480	109	0.0029		
Public bachelor's low education	3,839	21	0.0055	15,715	155	0.0099	241,254	740	0.0031		
Public master's high education	5,585	80	0.0143	31,732	1,143	0.0360	246,351	694	0.0028		
Public master's low education	33,677	175	0.0052	146,050	1,795	0.0123	1,206,699	3,636	0.0030		
Total public 4 year non-doctorate-granting	44,055	290	0.0066	196,563	3,206	0.0163	1,731,784	5,179	0.0030		
Public doctorate high education	8,243	97	0.0118	35,979	1,133	0.0315	261,796	719	0.0027		
Public doctorate low education	27,666	117	0.0042	125,234	1,380	0.0110	922,839	2,702	0.0029		
Public for-profit high education	12,317	148	0.0120	55,764	1,836	0.0329	427,319	1,175	0.0027		
Public for-profit low education	38,011	146	0.0038	227,253	2,531	0.0111	1,381,681	4,021	0.0029		
Total public 4 year non-doctorate-granting	86,237	508	0.0059	444,230	6,880	0.0155	2,993,635	8,617	0.0029		
Private not-for-profit less-than-2-year	†	†	†	†	†	†	22,462	601	0.0268		
Private not-for-profit 2 -year	†	†	†	†	†	†	115,789	1,201	0.0104		
Total private not-for-profit less-than-4-year	†	†	†	†	†	†	138,251	1,802	0.0130		
Private not-for-profit bachelor's high education	3,967	43	0.0108	13,208	380	0.0288	112,573	315	0.0028		
Private not-for-profit bachelor's low education	10,277	51	0.0050	46,082	532	0.0115	342,592	999	0.0029		
Private not-for-profit master's high education	7,929	96	0.0121	23,949	758	0.0317	196,492	543	0.0028		
Private not-for-profit master's low education	33,342	153	0.0046	79,456	896	0.0113	616,831	1,800	0.0029		
Total private not-for-profit 4 year non-doctorate-granting	55,515	343	0.0062	162,695	2,566	0.0158	1,268,488	3,657	0.0029		
Private not-for-profit doctorate high education	3,144	26	0.0083	9,314	237	0.0254	69,233	209	0.0030		
Private not-for-profit doctorate low education	4,792	24	0.0050	19,547	237	0.0121	133,560	418	0.0031		
Private not-for-profit first-professional high education	13,011	69	0.0053	52,907	889	0.0168	342,421	1,054	0.0031		
Private not-for-profit first-professional low education	8,975	68	0.0076	54,641	888	0.0163	281,436	856	0.0030		
Total private not-for-profit 4 year doctorate-granting	29,922	187	0.0062	136,409	2,251	0.0165	826,650	2,537	0.0031		
Private for-profit less-than-2-year	†	†	†	†	†	†	282,500	4,328	0.0153		
Private for-profit 2-year or more	6,235	38	0.0061	8,912	103	0.0116	553,660	1,823	0.0033		
Total private for-profit	6,235	38	0.0061	8,912	103	0.0116	836,160	6,151	0.0074		

Table G-2.—Student sampling rates used in determining measures of size by institutional stratum and type of student—Continued

Institutional stratum	GraduateStudents						First-professional students					
	Masters students			Doctoral students			Other graduate students					
	Size of universe	Sample size	Sampling rate	Size of universe	Sample size	Sampling rate	Size of universe	Sample size	Sampling rate			
Total	1,600,969	5,821	0.0036	362,237	4,543	0.0125	560,847	1,294	0.0023	317,626	1,319	0.0042
Public less-than-2-year	†	†	†	†	†	†	†	†	†	†	†	†
Public 2-year	†	†	†	†	†	†	†	†	†	†	†	†
Total less than 4-year	†	†	†	†	†	†	†	†	†	†	†	†
Public bachelor's high education	†	†	†	†	†	†	†	†	†	†	†	†
Public bachelor's low education	†	†	†	†	†	†	3,331	8	0.0024	†	†	†
Public master's high education	43,119	157	0.0036	†	†	†	21,722	51	0.0023	†	†	†
Public master's low education	214,383	783	0.0037	†	†	†	107,997	251	0.0023	†	†	†
Total public 4-year non-doctorate granting	257,502	940	0.0037	†	†	†	133,050	310	0.0023	†	†	†
Public doctorate high education	50,210	178	0.0035	16,877	206	0.0122	17,300	39	0.0023	†	†	†
Public doctorate low education	200,030	710	0.0035	67,237	821	0.0122	68,918	155	0.0022	†	†	†
Public first-professional high education	90,614	322	0.0036	30,458	371	0.0122	31,220	71	0.0023	15,732	63	0.0040
Public first-professional low education	329,334	1,169	0.0035	110,700	1,351	0.0122	113,468	255	0.0022	107,440	427	0.0040
Total public 4 year doctorate granting	670,188	2,379	0.0035	225,272	2,749	0.0122	230,906	520	0.0023	123,172	490	0.0040
Private not-for-profit 2-year or less	†	†	†	†	†	†	†	†	†	†	†	†
Private not-for-profit 2-year	†	†	†	†	†	†	†	†	†	†	†	†
Total private not-for-profit less than 4-year	†	†	†	†	†	†	†	†	†	†	†	†
Private not-for-profit bachelor's high education	†	†	†	†	†	†	†	†	†	†	†	†
Private not-for-profit bachelor's low education	47,472	168	0.0035	†	†	†	1,792	4	0.0022	†	†	†
Private not-for-profit master's high education	193,521	686	0.0035	†	†	†	13,002	29	0.0022	†	†	†
Private not-for-profit master's low education	240,993	854	0.0035	†	†	†	53,002	120	0.0023	†	†	†
Total Private not-for-profit 4-year non-doctorate-granting	35,332	134	0.0038	11,295	148	0.0131	67,796	153	0.0023	†	†	†
Private not-for-profit doctorate high education	†	†	†	†	†	†	11,294	27	0.0024	†	†	†
Private not-for-profit doctorate low education	71,957	274	0.0038	23,003	301	0.0131	23,002	56	0.0024	†	†	†
Private not-for-profit first-professional high education	113,361	432	0.0038	36,238	474	0.0131	36,238	88	0.0024	49,285	210	0.0043
Private not-for-profit first-professional low education	181,343	691	0.0038	57,970	758	0.0131	57,971	139	0.0024	143,499	612	0.0043
Total private not-for-profit 4 year doctorate-granting	401,993	1,531	0.0038	128,506	1,681	0.0131	128,505	310	0.0024	192,784	822	0.0043
Private for-profit less-than-2-year	†	†	†	†	†	†	†	†	†	†	†	†
Private for-profit 2-year or more	30,293	117	0.0039	8,459	113	0.0134	590	1	0.0017	1,670	7	0.0042
Total private for-profit	30,293	117	0.0039	8,459	113	0.0134	590	1	0.0017	1,670	7	0.0042

† Not applicable

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

and n_r^* is the number of non-certainty selections from stratum “r.” The sampling algorithm was implemented with a random start for each institutional stratum to ensure the positive pairwise probabilities of selection that are needed for proper variance estimation.⁴ The numbers of certainty and uncertainty schools selected, within each of the 22 institutional strata, are shown in table G-3.

Table G-3.—Institutional sampling rates and number of certainty and non-certainty institutions sampled, by institutional stratum

Institutional stratum ¹	Number of sample institutions		
	Total ²	Certainty	Noncertainty
Total	1,082	286	796
Public			
1 Less-than-2-year	34	8	26
2 2-year	198	9	189
Total less-than-4-year	232	17	215
3 Bachelor’s high education	5	0	5
4 Bachelor’s low education	19	1	18
5 Master’s high education	25	2	23
6 Master’s low education	78	6	72
Total 4-year non-doctorate-granting	127	9	118
7 Doctorate-granting high education	25	25	0
8 Doctorate-granting low education	63	31	32
9 First-professional-granting, high education	29	29	0
10 First-professional-granting low education	103	88	15
Total 4-year doctorate-granting	220	173	47
Private not-for-profit			
11 Less-than-2-year	12	0	12
12 2-year	23	2	21
Total less-than-4-year	35	2	33
13 Bachelor’s high education	17	0	17
14 Bachelor’s low education	37	0	37
15 Master’s high education	37	0	37
16 Master’s low education	82	6	76
Total 4-year non-doctorate-granting	173	6	167
17 Doctorate-granting, high education	16	7	9
18 Doctorate-granting, low education	27	4	23
19 First-professional-granting, high education	57	32	25
20 First-professional-granting, low education	68	34	34
Total 4-year doctorate-granting	168	77	91
Private, for-profit			
21 Less-than-2-year	77	0	77
22 2-year or more	50	2	48
Total private for-profit	127	2	125

¹Stratum reflects institutional categorization as determined from the 1998-99 IPEDS IC file; some errors in this classification were uncovered when institutions were contacted.

²During institutional contacting, we discovered that part of one school had recently split off and formed a separate institution. Both institutions were included in the sample, so the actual sample size is 1,083.

NOTE: “High education” refers to the 20 percent of institutions with the highest proportions of their baccalaureate degrees awarded in education (based on the 1996–97 IPEDS completions file). The remaining 80 percent of institutions were classified as “low education” (i.e., having a lower proportion of baccalaureate degrees awarded in education).

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

⁴J.R. Chromy, “Variance Estimators for a Sequential Sample Selection Procedure.” In *Current Top IMS in Survey Sampling*, ed. D. Krewski, R. Platek, and J.N.K. Rao (New York: Academic Press), 1981, 329-347.

Within each of the “r” institutional strata, additional implicit stratification was accomplished by sorting the sampling frame in a serpentine manner.⁵ For less-than-2-year, 2-year, and private, for-profit institutions the implicit strata were: (a) institutional level of offering (where levels had been collapsed to form strata); (b) the OBE Region from the IPEDS-IC file (Bureau of Economic Analysis of the U.S. Department of Commerce Region⁶); (c) FIPS state code; and (d) the institution measure of size. For public, 4-year and private not-for-profit, 4-year institutions, the implicit strata were: (a) Carnegie classifications or groupings of Carnegie classifications; (b) historically Black colleges and universities (HBCU) indicator; (c) the OBE Region from the IPEDS IC file; and (d) the institution measure of size. Table G-4 shows that the regional distribution of the sample is consistent with the sampling frame.

Table G-4.—Distribution of NPSAS:2000 institutional sample, by region

Region ¹	Sample institutions		IPEDS institutions ²	
	Number	Percent	Number	Percent
1. New England	70	6.5	394	6.1
2. Mid East	197	18.2	1,147	17.9
3. Great Lakes	163	15.1	945	14.7
4. Plains	85	7.9	584	9.1
5. Southeast	223	20.6	1,503	23.4
6. Southwest	104	9.6	623	9.7
7. Rocky Mountains	40	3.7	214	3.3
8. Far West	178	16.5	887	13.8
9. Outlying Areas	22	2.0	125	2.0

¹New England includes CT, ME, MA, NH, RI, VT; Mid East includes DE, DC, MD NJ, NY, PA; Great Lakes includes IL, IN, MI, OH, WI; Plains includes IA, KS, MN, MO, NE, ND, SD; Southeast includes AL, AR, FL, GA, KY, LA, MS, NC, SC, TN, VA, WV; Southwest includes AZ, NM, OK, TX; Rocky Mountains includes CO, ID, MT, UT, WY; Far West includes AK, CA, HI, NV, OR, WA; and Outlying Areas includes PR.

²Counts obtained from the sampling frame based on the 1998-99 IPEDS IC file.

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

⁵R.L. Williams and J.R. Chromy, “SAS Sample Selection MACROs.” *Proceedings of the fifth Annual SAS User’s Group International Conference*, 1980, 392-396.

⁶For sorting purposes, Alaska and Hawaii were put with Puerto Rico in the Outlying areas region rather than in the Far West region.

IV. The Student Samples

The initial student sample was selected from lists provided by 999 of the 1,072 institutions (from the original sample) that proved to be eligible. The postsecondary students eligible for NPSAS:2000 were those who attended a NPSAS-eligible institution during the previously defined NPSAS year 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) occupational or vocational program that required at least 3 months or 300 clock hours of instruction to receive a degree, certificate, or other formal award
- not concurrently enrolled in high school
- not enrolled *solely* in a GED or other high school completion program

A. Construction of Sampling Frames

Each of the 1,072 sampled institutions that were verified to be eligible for NPSAS:2000 was asked to provide lists of all its students who satisfied all the NPSAS eligibility conditions, preferably “unduplicated,” electronic lists (sent via e-mail, diskette, CD-ROM, or file transfer protocol (FTP)), together with identifying and classifying information (see Section 2.2.2 in the main report). Although electronic files were preferred, the preferences of sample institutions were accommodated, and whatever type(s) of student list(s) they were able to provide were accepted, as long as they were complete. (Final 1999-2000 enrollment lists were available from some institutions as early as January, 2000; however, other institutions could not provide final lists until December, 2000.) Separate, “unduplicated” lists (in which each student’s name appears only once) were requested for baccalaureate business, baccalaureate non-business, other undergraduate, masters, doctoral, other graduate, and first-professional students (the student sampling strata) from those institutions providing hard copy lists. As expected, however, many institutions sent separate lists for each term or course of instruction; in which cases an individual student’s name could appear on more than one list. In such cases, procedures were used to “unduplicate” the sample, to ensure that each student received only one chance of selection.

B. Student Sample Selection

Students were sampled on a flow basis as student lists were received. Stratified systematic sampling was used to ensure comparable sampling procedures for both hard-copy and electronic lists. For each institution, the student sampling rates, rather than the student sample sizes, were set to fixed values:

- to facilitate sampling students on a flow basis as student lists were received
- to facilitate the procedures used to “unduplicate” the samples selected from (duplicated) hard-copy lists

- because sampling at a fixed rate based on the overall stratum sampling rates and the institutional probabilities of selection results in approximately equal overall probabilities of selection within the ultimate institution-by-student strata

Electronic lists were “unduplicated” by sorting on the student ID number and deleting duplicates prior to sample selection. In the case of duplicated hard-copy lists, a stratified systematic sample was selected from each list provided (typically separate lists by term). For unduplication, if there was a separate baccalaureate list, all students in the sample selected from the baccalaureate list were retained for the sample, and the *samples* selected from all other lists were “unduplicated” against the baccalaureate list. The baccalaureate list was given precedence since a student receiving a bachelors degree was sampled as a baccalaureate regardless of student type. After giving precedence to baccalaureates or if there was not a separate baccalaureate list, non-baccalaureate students in the sample selected from the fall list were retained for the sample, and the *samples* selected from all other lists were “unduplicated” against the fall list. (The fall term was given precedence in this process for comparability with NPSAS:87, in which only fall enrollees were sampled.) If the institution did not have standard terms, other orderings of the student lists were used to achieve unduplication of the sample.

After the sample of students had been selected for an institution, the social security numbers of the sample students were compared to those of students who had already been selected from other institutions. When duplicates were detected, the duplicate was eliminated from the sample from the current institution so that no student would be included in the sample twice. Multiplicity adjustments in the sample weighting (see Chapter 6) account for the fact that any students who attended more than one institution in the NPSAS population had more than one chance of selection.

The development of student sampling rates within student stratum “rs” (i.e., the r-th institutional stratum and the s-th student stratum within institutional stratum) were previously discussed in Section III.B, and the notation used in that development will be used here. For graduate and first-professional students, these overall student sampling rates were shown in table G-2.

For the unconditional probability of selection to be a constant for all eligible students in stratum “rs,” the overall probability of selection should be the overall student sampling fraction, f_{rs} ; i.e., we must require that

$$\frac{m_{rs}(j)}{M_{rs}(j)} \pi_r(j) = f_{rs} ,$$

or equivalently,

$$m_{rs}(j) = f_{rs} \frac{M_{rs}(j)}{\pi_r(j)} .$$

Thus, the conditional sampling rate for stratum “rs,” given selection of the j-th institution, becomes

$$f_{rs|j} = f_{rs} / \pi_r(j) .$$

It should be noted that, in this case, the desired overall student sample size, m_s , is achieved only

in expectation over all possible samples.

Achieving the desired sample sizes with equal probabilities within strata in the particular sample that has been selected and simultaneously adjusting for institutional nonresponse and ineligibility requires that

$$\sum_{j \in R} m_{rs}(j) = m_{rs},$$

where “R” denotes the set of eligible, *responding* institutions. Letting the conditional student sampling rate for stratum “rs” in the j-th institution be

$$\hat{f}_{rs|j} = \hat{f}_{rs} / \pi_r(j),$$

then requires

$$\sum_{i \in R} \hat{f}_{rs} \frac{M_{rs}(j)}{\pi_r(j)} = m_{rs},$$

or equivalently,

$$\hat{f}_{rs} = m_{rs} / \hat{M}_{rs}$$

where

$$\hat{M}_{rs} = \frac{\sum_{j \in R} M_{rs}(j)}{\pi_r(j)}.$$

Since it was necessary to set student sampling rates before complete information on institutional eligibility and response status was available, \hat{M}_{rs} was calculated as follows:

$$\hat{M}_{rs} = \sum_{j \in S} \frac{M_{rs}(j)}{\pi_r(j)} * E_r * R_r * E_{rs},$$

where “S” denotes the set of all sample institutions,

E_r	=	the institutional eligibility factor for institutional stratum “r”
R_r	=	the institutional response factor for institutional stratum “r”
E_{rs}	=	the student eligibility factor for student stratum “rs”

These factors were the proportions of institutions or students, respectively, expected to be eligible or responding within the defined strata. Since this determination was made after eligibility status had already been determined for some institutions, values of 0 (known not eligible) or 1 (known eligible) were used, if known at that time.

Initial student sampling rates were calculated in this manner for each sample institution; these rates were designed to achieve equal probabilities of selection within the ultimate institution-by-student sampling strata. However, these rates were sometimes modified for reasons listed below.

- The student sampling rates were ratio adjusted upwards, as needed based on enrollment file counts, so that the sample size achieved at each sample institution would be at least 40 sample students, where possible. (The reason for this constraint was to facilitate obtaining at least 30 responding students for most participating

institutions. The student sampling rates were decreased if the sample size was more than 50 greater than the institution had been told to expect, which was based on the sampling rate applied to the enrollment count, on the sampling frame. (This was to facilitate continued participation by the institutions for CADE data abstraction).

- The sample yield was monitored throughout the months during which student lists were received, and the student sampling rates were adjusted periodically for institutions for which sample selection had not yet been performed to ensure that the desired student sample sizes were achieved.

These adjustments to the initial sampling rates (especially the first two types of adjustments) resulted in some additional variability in the student sampling rates, and, hence, in some increase in survey design effects.

The planned and achieved sample sizes by student stratum and level of offering are shown in table G-5. Table G-5 shows that the rate adjustment procedures were generally effective; the overall sample yield was very close to what was planned (70,232 students as compared to the target of 70,266). The actual sample sizes achieved in total and within each institutional and student stratum, are shown in table G-6.

Table G-5.—Planned and achieved NPSAS:2000 student samples, by student stratum and level of offering

Student stratum ¹	Institutional level ²	Students sampled		
		Number expected ³	Number achieved	Percent ⁴
Total	All institutions	70,266	70,232	100.0
Baccalaureate business	4-year	1,365	1,475	108.1
Baccalaureate other	4-year	15,006	15,147	100.9
Other undergraduate	All	40,918	40,981	100.2
	Less-than-2-year	6,925	6,665	96.2
	2-3 Year	12,653	13,240	104.6
	4+ Year	21,340	21,076	98.8
Master's	4-year	5,820	5,964	102.5
Doctorate	4-year	4,543	3,946	86.9
Other graduate	4-year	1,293	1,369	105.9
First-professional	4-year	1,319	1,350	102.4

¹As expected, the sampling frames misclassified some individual students as to baccalaureate, undergraduate, graduate, and first-professional status; statistics presented in this table are based on the sampling frame classification.

²Institutional level is based on level confirmed by institution during school contacting.

³Based on sample allocation, 1998–1999 IPEDS IC file enrollment counts, and 1996–1997 IPEDS Completions file baccalaureate counts.

⁴Percent reported reflects the ratio of "achieved" to "planned."

⁵A percentage of each institution's graduate students were expected to be other graduate students (such as non-degree graduate or post-baccalaureate students) depending on type of institution, however the actual percentage of other graduate students varied by institution.

NOTE: Numbers may not sum to total due to rounding.

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

Table G-6.—Initial classification of NPSAS:2000 student sample, by type of institution and student stratum

Institution type	Student sampling stratum ¹												
	Total sample		Baccalaureate sample ²			Other undergraduate sample			Graduate sample ²			First-professional sample	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	
All institutions	70,232	100.0	16,622	100.0	40,981	100.0	11,279	100.0	1,350	100.0			
Institutional level													
Less-than-2-year	6,665	9.5	†	†	6,665	16.3	†	†	†	†			
2-year	13,240	18.9	†	†	13,240	32.3	†	†	†	†			
4-year non-doctorate granting	18,754	26.7	6,645	40.0	9,824	24.0	2,285	20.3	†	†			
4-year doctorate granting	31,573	45.0	9,977	60.0	11,252	27.5	8,994	79.7	1,350	100.0			
Institutional control													
Public	43,748	62.3	10,745	64.6	25,974	63.4	6,537	58.0	492	36.4			
Private not-for-profit	19,372	27.6	5,629	33.9	8,472	20.7	4,413	39.1	858	63.6			
Private for-profit	7,112	10.1	248	1.5	6,535	16.0	329	2.9	†	†			
Institutional sector													
Public less-than-2-year	1,527	2.2	†	†	1,527	3.7	†	†	†	†			
Public 2-year	10,663	15.2	†	†	10,663	26.0	†	†	†	†			
Public 4-year non-doctorate granting	9,884	14.1	3,464	20.8	5,208	12.7	1,212	10.8	NA	†			
Public 4-year doctorate granting	21,674	30.9	7,281	43.8	8,576	20.9	5,325	47.2	492	36.4			
Private not-for-profit 2-year or less	1,836	2.6	†	†	1,836	4.5	†	†	†	†			
Private not-for-profit 4-year non-doctorate granting	8,005	11.4	3,033	18.3	4,043	9.9	929	8.2	†	†			
Private not-for-profit 4-year doctorate granting	9,531	13.6	2,596	15.6	2,593	6.3	3,484	30.9	858	63.6			
Private for-profit less-than-2-year	4,523	6.4	†	†	4,523	11.0	†	†	†	†			
Private for-profit 2-year or more	2,589	3.7	248	1.5	2,012	4.9	329	2.9	†	†			

† Not applicable.

¹As expected the sampling frames misclassified some individual students as to baccalaureate undergraduate graduate and first-professional status; statistics presented in this table are based on the sampling frame classification.

²For this presentation the two baccalaureate strata have been combined and the masters, doctorate, and other graduate strata have been combined. SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000).

