

Chapter 2

Design and Method of NPSAS:2000

Implementation of NPSAS:2000 required a complex set of study tasks and activities, including sequentially dependent data collection operations as well as overlapping development, analysis, documentation and reporting tasks. An RTI-developed Integrated Management System (IMS), used effectively in other large-scale survey projects, was adapted, based on results of an extensive NPSAS field test,¹ for use in the full-scale NPSAS:2000.

2.1 NPSAS:2000 Target Population and Sampling Overview

The basic features of the NPSAS:2000 sampling plan and the resulting samples are summarized in the sections 2.1.1 and 2.1.2. Greater detail is provided in appendix G for the interested reader.

2.1.1 Target Population

The target population for NPSAS:2000 consisted of all students enrolled at any time in postsecondary institutions in the United States or Puerto Rico and which had signed Title IV participation agreements with the U.S. Department of Education making them eligible for the federal student aid programs (Title IV institutions) between July 1, 1999, and June 30, 2000.² 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. The exception occurred 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.³ This definition of the survey population differed from previous NPSAS rounds but was more consistent with the definition of the target population. More specific definitions of the institution and student populations are provided in section 2.2.

Though NPSAS:2000 was limited to Title IV institutions, prior NPSAS rounds also surveyed students enrolled at institutions not participating in Title IV aid programs. In addition,

¹U.S. Department of Education. National Center for Education Statistics. *National Postsecondary Student Aid Study (NPSAS:2000) Field Test Methodology Report*, NCES No. 2000-17, by Melissa R. Biber, Michael W. Link, John A. Riccobono, and Peter H. Siegel. Andrew G. Malizio, project officer. Washington, DC: October 2000.

² Excluding students who were enrolled in military service academies, were enrolled solely in a General Education Development (GED) program, or were concurrently enrolled in high school.

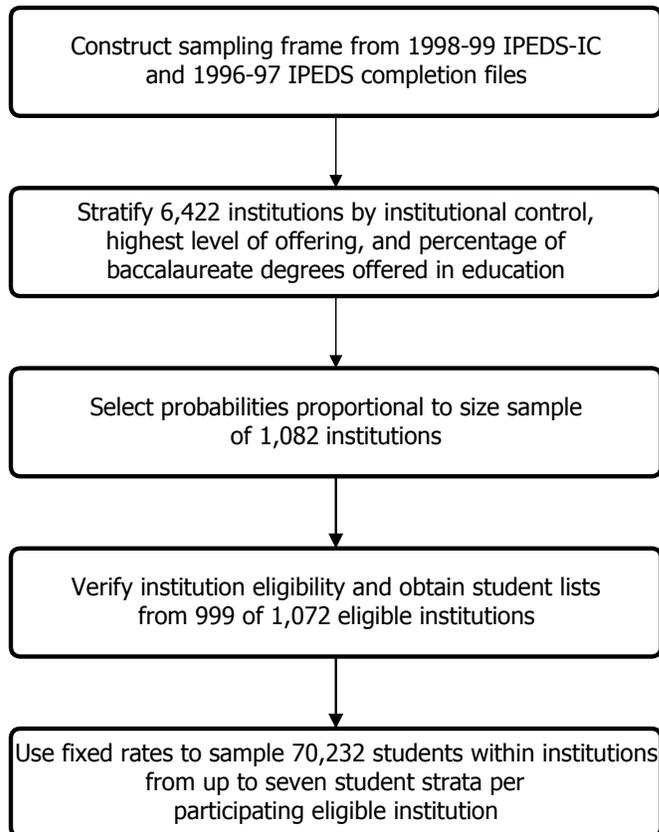
³ The target population is the population about which inferences will be made. The survey population is the population actually covered by the sampling frame. Nearly all members of the target population were also members of the survey population; however, the adopted definition of the survey population 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). Poststratification adjustments of the analysis weights (see Chapter 6) reduce any resulting bias for inferences regarding the target population.

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.

2.1.2 Sample Design Overview

An overview of the sequential statistical sampling process for NPSAS:2000 is provided in figure 2-1. The institutional sampling frame for NPSAS:2000 was constructed from the 1998–99 Integrated Postsecondary Education Data System Institutional Characteristics (IPEDS-IC) file and, because NPSAS:2000 also served as the base-year survey for a Baccalaureate and Beyond longitudinal study, the 1996-97 IPEDS completion file was used to check frame information regarding estimated size of institutional graduating classes. Both of these files were the latest available at the time of NPSAS institutional sampling.

Figure 2-1.—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 IPEDS-IC database provided nearly complete coverage of the institutions in the target population. Listings in the file that were not eligible institutions (e.g., institutions located outside the U.S. and Puerto Rico; central offices; military academies) were deleted from the population file. Additional information for eligible institutions was obtained from the 1996–97 IPEDS completion files. The eligible institutions 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 bachelor's high education⁴
4. Public bachelor's low education
5. Public master's high education
6. Public master's low education
7. Public doctorate-granting high education⁵
8. Public doctorate-granting low education
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 bachelor's high education
14. Private not-for-profit bachelor's low education
15. Private not-for-profit master's high education
16. Private not-for-profit master's 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. The sampling frames for selecting sample students were paper-copy and electronic lists of students provided by the sample institutions for those students enrolled in terms or courses of instruction during the previously defined NPSAS year.⁶ Student lists were sampled on a flow basis as they were received, using stratified systematic sampling. The seven student sampling strata were as follows:

⁴ For each category that had a high education and low education breakout, the high education stratum was defined to be 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 constituted the low 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 longitudinal study.

⁵ Institutions that awarded first-professional degrees were included in the doctorate-granting stratum.

⁶ Quality control checks were performed on each list received from a sample institution, by comparing the numbers of undergraduate, graduate, and first-professional students listed to the “unduplicated” head counts reported for the 1997–98 academic year in the 1998–99 IPEDS-IC file. The number of baccalaureates listed was compared to the counts reported for the 1996–97 academic year in the 1996–97 IPEDS completions file.

1. Students receiving a baccalaureate degree in business⁷
2. Other baccalaureate recipients
3. Other undergraduate students
4. Master's students
5. Doctoral students
6. Other graduate students
7. First-professional students

The list for each student stratum was sampled at a rate designed to provide approximately equal student-level probabilities. Student sampling rates were revised after enough lists had been received to more accurately estimate the overall sample yield. These sampling procedures resulted in selection of 70,232 students.

2.2 NPSAS:2000 Sample Implementation

The goal of all sampling activities was to attain the targeted numbers of eligible sample postsecondary students within each of the specified student and institution strata. An important domain of the student sample was the set of students identified as baccalaureates,⁸ who are the baseline cohort for the Baccalaureate and Beyond (B&B) longitudinal study. The desired number of sample students was determined by accounting for expected (from prior NPSAS rounds) rates of nonresponse and ineligibility among sample students and rates of B&B misclassification (as determined from NPSAS:93 and the NPSAS:2000 field test). Since the student samples were selected 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. For each institution, these rates were set based on the institution's probability of selection and the overall student stratum sampling rates. The sampling rates were set to meet or exceed the sample sizes shown in table 2-1.

The NPSAS:2000 sample was also designed to obtain at least 30 student CATI respondents from each sample institution that had at least that many eligible students enrolled during the NPSAS year. 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 1,008.⁹ Based on institutional participation rates obtained in prior NPSAS rounds and the NPSAS:2000 field test, an initial sample of 1,082 institutions was selected.

⁷ Students receiving a baccalaureate degree in business were in a separate stratum so that they would be selected at a lower sampling rate than other baccalaureate recipients, because sampling them at the same rate would result in more students receiving a baccalaureate degree in business than desired.

⁸ Students who received their bachelor's degree during the 1999–2000 academic year.

⁹ An institution was considered participating if it sent in a usable enrollment list.

Table 2-1.—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	236	127	109	†	†
4 Bachelor's low education	923	175	740	†	†
5 Master's high education	2,124	1,223	694	208	†
6 Master's 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 Bachelor's high education	739	423	315	†	†
14 Bachelor's low education	1,586	583	999	†	†
15 Master's high education	1,595	855	543	197	†
16 Master's 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.

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).

2.2.1 Institutional Sample

The target population for NPSAS:2000 included nearly all Title IV participating postsecondary institutions in the 50 states, the District of Columbia, and Puerto Rico.¹⁰

¹⁰ Title IV participating institutions excluded from the target population were the five U.S. service academies.

To be eligible for NPSAS:2000, an institution was required, during the 1999–2000 academic year, to:¹¹

- offer an educational program designed for persons who had 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 were open to more than the employees or members of the company or group (e.g., union) that administered the institution;
- be located in the 50 states, the District of Columbia, or Puerto Rico;
- be other than a U.S. Service Academy;¹² and
- have a signed Title IV participation agreement with the U.S. Department of Education.

As indicated above, institutions providing only avocational, recreational, or remedial courses or only in-house courses for their own employees were excluded.

The student sample was allocated to the separate applicable institutional and student sampling strata, defined above. Student sampling rates, which were used to compute institution-level composite measures of size, were based on 1998–99 IPEDS IC and 1996–97 IPEDS completions file counts and the required sample sizes (see appendix G for details).

An independent sample of institutions was selected for each institutional stratum using Chromy's¹³ sequential probability minimum replacement (pmr) sampling algorithm to select institutions with probabilities proportional to their computed measures of size. However, rather than multiple selections of sample institutions being allowed,¹⁴ those with expected frequencies of selection greater than unity (1.00) were selected with certainty. The remainder of the institutional sample was selected from the remaining institutions within each stratum. The sampling algorithm was implemented with a random start for each institutional stratum to ensure the positive pairwise probabilities of selection that were needed for proper variance estimation.¹⁵

¹¹The listed eligibility requirements are consistent with those used in previous NPSAS rounds, except for the last one.

¹²These academies were not eligible for this financial aid study because of their unique funding/tuition base.

¹³J.R. Chromy. "Sequential Sample Selection Methods." *Proceedings of the American Statistical Association Section on Survey Research Methods of the American Statistical Association*, 1979, 401–406.

¹⁴Precluding institutions with multiple selections at the first stage of sampling made it unnecessary to select multiple second-stage samples of students.

¹⁵J.R. Chromy (1981). Variance Estimators for a Sequential Sample Selection Procedure. In D. Krewski, R. Platek, and J.N.K. Rao (Eds.), *Current Top IMS in Survey Sampling* (pp. 329-347). New York: Academic Press.

The numbers of certainty and noncertainty schools selected, within each of the 22 institutional strata, are shown in table 2-2. Within each institutional stratum, additional implicit stratification was accomplished by sorting the stratum sampling frame in a serpentine manner.¹⁶ For less-than-2-year, 2-year, and private for-profit institutions, the implicit strata were: (1) institutional level of offering (where levels had been collapsed to form strata); (2) the OBE Region from the IPEDS IC file (Bureau of Economic Analysis of the U.S. Department of Commerce Region);¹⁷ (3) the Federal Information Processing Standard (FIPS) state code; and (4) the institution measure of size. For public 4-year and private not-for-profit 4-year institutions, the implicit strata were: (1) Carnegie classifications of postsecondary institutions or groupings of Carnegie classifications; (2) historically black colleges and universities (HBCU) indicator; (3) the Region from the IPEDS-IC file; and (4) the institution measure of size. The objectives of this additional, implicit stratification were to approximate proportional representation of institutions on these measures. Table 2-3 shows that the regional distribution of the sample is consistent with the sampling frame.

2.2.2 Student Sample

The postsecondary students eligible for NPSAS:2000 were those who attended a NPSAS-eligible institution during the 1999–2000 academic year and who were

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

Each sampled institution that was verified as NPSAS-eligible was asked to provide lists of all its students who satisfied all the NPSAS eligibility conditions, preferably “unduplicated” (i.e., each student’s name appeared only once) electronic lists (sent via e-mail, diskette, CD-ROM, or file transfer protocol [FTP]), together with identifying and classifying information (see Section 2.2.3. below). 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. Separate, “unduplicated” lists were requested for baccalaureate business, baccalaureate nonbusiness, other undergraduate (i.e., non-baccalaureate undergraduates), master’s, doctoral, other graduate, and first-professional students (the sampling strata) from institutions providing paper-copy lists. As expected, however, many institutions

¹⁶ R.L. Williams, and J.R. Chromy. “SAS Sample Selection MACROS.” *Proceedings of the Fifth Annual SAS Users Group International Conference*, 1980, 392–396.

¹⁷ For sorting purposes, Alaska and Hawaii were combined with Puerto Rico in the Outlying Areas region rather than in the Far West region.

Table 2-2.—Institutional sampling rates and number of certainty and noncertainty institutions sampled, by institutional stratum

Institutional stratum ¹	Size of universe ²	Sampling Rate	Number of sample institutions		
			Total ³	Certainty	Noncertainty
Total	6,422	0.17	1,082	286	796
Public					
1 Less-than-2-year	255	0.14	34	8	26
2 2-year	1,208	0.16	198	9	189
Total less-than-4-year	1,463	0.16	232	17	215
3 Bachelor's high education	18	0.29	5	0	5
4 Bachelor's low education	69	0.27	19	1	18
5 Master's high education	51	0.49	25	2	23
6 Master's low education	196	0.40	78	6	72
Total 4-year non-doctorate-granting	334	0.38	127	9	118
7 Doctorate-granting high education	25	1.00	25	25	0
8 Doctorate-granting low education	82	0.77	63	31	32
9 First-professional-granting high education	29	1.00	29	29	0
10 First-professional-granting low education	115	0.89	103	88	15
Total 4-year doctorate-granting	251	0.88	220	173	47
Private not-for-profit					
11 Less-than-2-year	112	0.10	12	0	12
12 2-year	314	0.07	23	2	21
Total less-than-4-year	426	0.08	35	2	33
13 Bachelor's high education	112	0.15	17	0	17
14 Bachelor's low education	402	0.09	37	0	37
15 Master's high education	120	0.31	37	0	37
16 Master's low education	414	0.20	82	6	76
Total 4-year non-doctorate-granting	1,048	0.16	173	6	167
17 Doctorate-granting high education	24	0.66	16	7	9
18 Doctorate-granting low education	88	0.31	27	4	23
19 First-professional-granting high education	80	0.71	57	32	25
20 First-professional-granting low education	294	0.23	68	34	34
Total 4-year doctorate-granting	486	0.35	168	77	91
Private for-profit					
21 Less-than-2-year	1,386	0.06	77	0	77
22 2-year or more	1,028	0.05	50	2	48
Total private for-profit	2,414	0.05	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.

²Based on the 1998–99 IPEDS IC file.

³During institutional contacting, it was discovered that part of one school had recently split off and formed a separate institution. Both institutions were included in the sample, adding another institution to stratum 10, so the actual total 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).

Table 2-3.—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.

NOTE: Details may not sum to 100 due to rounding.

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

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, the samples were “unduplicated” to ensure that each student received only one chance of selection.¹⁸

As student lists were received from institutions, students were sampled. Stratified systematic sampling was used to ensure comparable sampling procedures for both paper-copy and electronic lists. In the case of duplicated paper-copy lists, a stratified systematic sample was selected from each list provided (typically separate lists by term) and the samples selected were “unduplicated” against master lists (see appendix G).¹⁹ After the sample of students had been selected for an institution, Social Security numbers (SSNs) of those sampled were compared to those of students who had already been selected from other institutions to eliminate cross-institution duplication. Multiplicity adjustments in the sample weighting described in more detail in Chapter 6 accounted for the fact that any students who attended more than one institution during the NPSAS year had more than one chance of selection.

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

¹⁸ Electronic lists were “unduplicated” by sorting on the student identification (ID) number and deleting duplicates prior to sample selection.

¹⁹ The baccalaureates were given precedence since a student receiving a bachelor's degree was sampled as a baccalaureate regardless of student type. Next, the fall term was given precedence in this process for comparability with NPSAS:87. If the institution did not have standard terms, other orderings of the student lists were used to achieve unduplication of the sample.

2. Design and Method of NPSAS:2000

- The student sampling rates were increased, as needed, so that the sample size achieved at each sample institution would be at least 40 sample students, where possible.
- 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.²⁰
- 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 (variance inflation—see Chapter 6).

The planned and achieved sample sizes by student stratum and level of offering are shown in table 2-4. The actual sample sizes achieved in total and by school type and student stratum are shown in table 2-5. Table 2-4 shows that the overall sample yield was very close to what was planned (70,232 students as compared to the target of 70,266). This table also shows that overall there were more baccalaureate, master's, other graduate, and first-professional students in the sample than planned, and there were fewer doctoral students than planned.

Table 2-4.—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 ⁴	Percent ⁵
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	Subtotal	40,918	40,981	100.2
	Less-than-2-year	6,925	6,665	96.2
	2- to 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
Doctor's	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, 1998–99 IPEDS IC file enrollment counts, and 1996–97 IPEDS completions file baccalaureate counts. Numbers may not sum to total due to rounding.

⁴The student sample was drawn from 999 institutions determined to be eligible and providing enrollment lists.

⁵Percent reported reflects the ratio of “achieved” to “expected.”

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

²⁰ This was to facilitate continued participation by the institutions for CADE data abstraction.

Table 2-5.—Initial classification of NPSAS:2000 student sample, by type of institution and student stratum

Institution type	Total sample ¹		Student sampling stratum ²							
	Number	Percent	Baccalaureate sample ³		Other undergraduate sample		Graduate sample ³		First-professional sample	
			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	†	†
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.

¹The student sample was drawn from 999 institutions determined to be eligible and providing enrollment lists.

²As 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.

³The two baccalaureate strata have been combined and the master's, doctorate, and other graduate strata have been combined.

NOTE: Details may not sum to 100 due to rounding.

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

2.2.3 Institutional Enlistment and Student List Acquisition and Sampling

Once institutions were sampled, attempts were made to contact the chief administrator of the selected institutions to verify institutional eligibility, solicit participation of eligible institutions, and request appointment of an Institutional Coordinator through which subsequent communication with the institution would be directed. The initial letter on U.S. Department of Education (ED) letterhead included a study fact sheet and endorsement letters, as appropriate for that institution, from the National Association of Financial Aid Administrators (NASFAA), the American Association of College Registrars and Admissions Officers (AACRAO), the Career College Association (CCA), and the National Accrediting Commission of Cosmetology Arts and Sciences (NACCAS). Concurrently, NASFAA mailed a separate letter directly to the financial aid officers of all member institutions sampled urging participation. (Copies of these letters and attachments, as well as other correspondence mailed to sampled institutions or students during the course of the full-scale survey are included in appendix B.) Follow-up telephone calls were made to the chief administrator one week after the mailing; if the IC had not been named by that time, the administrator was urged to name an Institutional Coordinator (with varying degrees of success) during the telephone conversation.

Separate mailings to the Institutional Coordinators (containing all materials included in the initial mailing to the chief administrator) were initiated on a flow basis, as the Institutional Coordinators were designated. Follow-up telephone calls were, again, initiated one week following the mailing (the initial contact with the Institutional Coordinators typically involved a series of calls, including refusal conversion calls, since no substitution of refusing institutions was employed). Institutional coordinators were advised of what would be expected from the institution and asked to verify the IPEDS classification (institutional control and highest level of offering) and the calendar system used (including dates that terms started). Institutional Coordinators also were asked to (1) provide information on the institution's record-keeping approaches (including identifying the physical on-campus locations of records needed for the subsequent record abstraction procedures), (2) identify their PC capabilities for operating the CADE software, and (3) set a date by which the school would provide student enrollment lists.

The list(s) requested (preferably a single "unduplicated" electronic list) were to contain all eligible students enrolled in any term within the study-defined year. (Sampled schools with additional NPSAS-year terms starting after the date of the request obviously could not provide complete lists until after the last applicable term began.) The data items requested for each listed student were

- full name;
- student identification (ID) number;
- Social Security number (possibly identical with student ID);
- educational level—undergraduate, master's, doctoral, other graduate, or first professional—during the *last* term of enrollment during the study-defined year,
- for baccalaureate students major field of study for which the baccalaureate degree was or will be awarded; and

- Classification of Instruction Program (CIP) code for the student's major.

Definitions of types of lists and information preferred, as well as instructions for preparing different lists, were included in the initial IC letter and further clarified, as needed, in follow-up telephone conversations. In such subsequent telephone contacts, contractor staff worked closely with the IC to determine the best reasonable alternative lists and student information that could be provided by the institution.

Prompting telephone calls were made to institutions that had not provided lists by one week following the most recent delivery date previously agreed upon by the IC. Throughout the list acquisition process, attempts were made by the contractor to accommodate school constraints and to reduce their burden, including contractor "unduplication" of lists. Where requested, institutions were reimbursed for personnel and computer time required to prepare student sampling lists.

Several checks on quality and completeness of student lists were implemented before the sample students were selected. Institutions providing lists that failed these checks were called to rectify the detected problems. Completeness checks were failed if any of the following conditions existed:

- Baccalaureate recipients/graduating seniors were not identified (unless the institution was less-than-4-years or explicitly indicated that no such students existed in the school).
- Student level—undergraduate, master's, doctoral, other graduate, or first professional—was not clearly identified.
- Major fields of study or CIP codes were not clearly identified for baccalaureates.

Quality checks were performed by checking the "unduplicated" count from provided lists against the "unduplicated" counts from IPEDS and completions files. For applicable institutions, separate checks were made for baccalaureate recipients, other undergraduates, graduate, and first-professional students; for institutions serving only undergraduates (and no baccalaureates), checks were made against total enrollment. The institution failed the check if the count for any "unduplicated" list differed by at least 25 percent from the IPEDS count.²¹

2.3 Data Collection and Operational Design

NPSAS:2000 involved a multistage effort to collect information related to student aid. An initial NPSAS:2000 data collection stage collected electronic student aid report (Institutional Student Information Report, or ISIR) information directly from the U.S. Department of Education Central Processing System (CPS) for federal financial aid applications.²² The second

²¹ If provided lists were not "unduplicated," the contractor estimated the "unduplicated" total by applying an empirically determined multiplicity factor (0.50) to the count over provided lists; in these cases, the critical difference also was relaxed to at least 30 percent.

²² The contractor for this service was National Computer Systems (NCS). Students completed a Free Application for Federal Student Aid (FAFSA), which was mailed to the CPS contractor; this information was entered into the computer file and electronic versions of the Institutional Student Information Record (ISIR) were created. The ISIR information was made available to all institutions that the student indicated on the FAFSA.

stage involved abstracting information from the student's records at the school from which he/she was sampled, using a computer-assisted data entry (CADE) system. In the third stage, interviews were conducted with sampled students, primarily using a computer-assisted telephone interviewing (CATI) procedure. Computer-assisted personal interviewing (CAPI) procedures, using field interviewers, were also used for the first time on a NPSAS study, to help reduce the level of nonresponse to CATI.

A schematic of the operational flow of major data collection components of the NPSAS:2000 study is shown in figure 2-2 and discussed below. To meet established dates for conclusion of all activities, while accommodating both differential dates at which student sampling could be initiated and differential timeliness of institutional turnaround, not all stages were implemented at the same time at all institutions. In fact, the only fixed points in operations were (1) selection of the institutional sample plus the initial institutional mailings and verification calls, and (2) cutoff of interviewing. Start and end dates for the significant study activities were shown earlier in table 1-1.

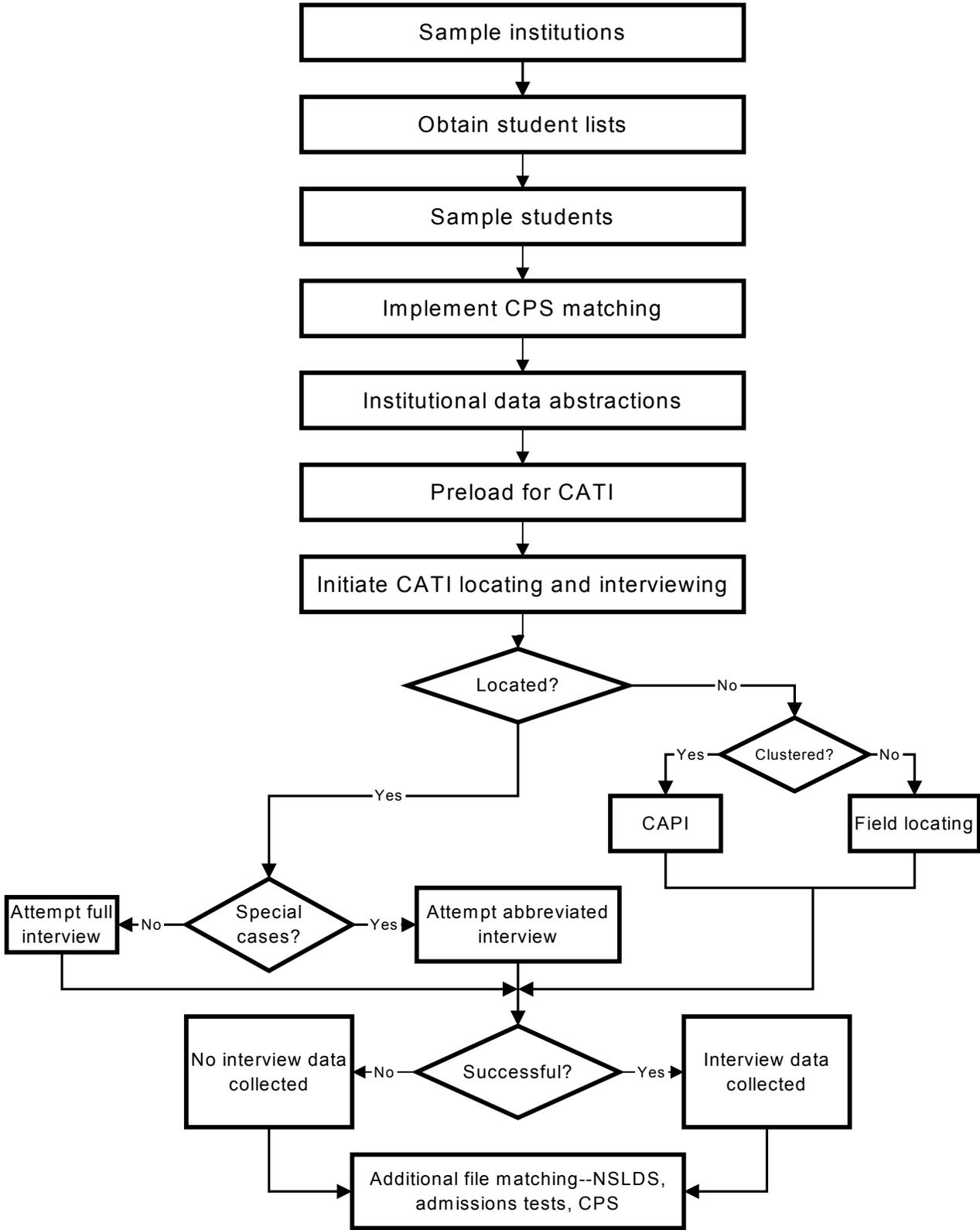
2.3.1 Overview of Data Collection Instruments and Extant Data Sources

As noted previously, some study data were obtained from extant databases. These additional data sources served several useful functions. First, they provided information that could not be collected from the institutions or the students. Second, they provided a way to "fill in" data that was obtained in institutional record abstraction or the student interview but was missing for individual sample members (e.g., demographics). Also, additional data sources served as a way to check or confirm information obtained from student records or interviews.

Information related to applications for federal financial aid was obtained (for two academic years) from ED's central processing system, the CPS. Additionally, data on the nature and amounts of received Pell grant or federal student loans were obtained from the National Student Loan Data System (NSLDS) databases maintained by ED. The NSLDS Pell grant and loan files that were accessed included information for the 1999–2000 academic year as well as a complete federal grant or loan history for each applicable student. In addition to information regarding student aid receipt, data were obtained from Educational Testing Service (ETS) for the SAT, and from ACT for the ACT assessment, which included test score data as well as additional demographic information and some information regarding educational aspirations.

Obtaining Central Processing System (CPS) information. To reduce institutional burden in subsequent data collections, the NPSAS:2000 contractor, with the assistance of NCES, arranged to obtain information from the Central Processing System (which was operated for the U.S. Department of Education by a separate contractor, National Computer Systems [NCS]), to access certain information provided by all federal financial aid applicants who had been selected in the sample. Students give this information to the CPS contractor on a Free Application for Federal Student Aid (FAFSA) form; it is then converted to electronic form, analyzed, and provided to involved schools (and other approved parties).

Figure 2-2.—Flow of major data collection components for the NPSAS:2000 study



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

CADE data abstraction from students' institutional records. Data from sampled students' records at the NPSAS institution were collected using procedures similar to those successfully tested and implemented during NPSAS:96. Specifically, a CADE software system using version 4.3 of the Computer Assisted Survey Execution System (CASES)²³, was developed for use in collecting data from student records. The data elements included in the Web-CADE system (described in more detail in chapter 3) were identical to those included in the laptop-based CADE system used by the RTI field data collectors (field-CADE).

The CADE record abstraction process began when a student sample had been selected and transmitted to the Central Processing System for obtaining financial aid application data. Upon completion of the CPS matching (typically a 48-hour turnaround), a number of data elements were preloaded into the CADE database, thus initializing the CADE system. These preloaded elements included an indicator of whether the student had been matched successfully to the CPS system, as well as selected CPS variables for use in CADE software edit checks. In addition, the system was customized for each institution by preloading of the names of up to 10 institution financial aid programs and up to 10 state financial aid programs, for use in identifying aid received by students.

As was the case in NPSAS:96, institutions could choose either to enter the data themselves or to have an RTI-employed field data collector enter the data. Institutions were encouraged to use their own staff for this data collection (with compensation for staff time, when requested), since this minimized the overall cost of the data collection. The NPSAS:2000 field test demonstrated the effectiveness and user-friendliness of the Web-CADE system, providing institutions with further encouragement to complete the data collection themselves.

Once CADE was initialized for a particular institution, the Institutional Coordinator was notified by telephone that the CADE data collection could begin. Coordinators who had previously indicated a willingness to complete the data collection via Web-CADE were provided with a user name and password to gain access to the Web-CADE systems. As a security measure, each coordinator was asked to provide a "lost-password prompting question and answer"—that is, if they forgot their password and had to call in for a reminder, the personalized question was posed and the password was provided when they successfully answered the question. Field-CADE institutions were also notified by telephone of CADE initialization, at which time an appointment was made for a field data collector to visit the institution.

The CADE software (the full contents of which appear in appendix E) was structured into eight sections:

1. locating – for collecting address and phone information for students, students' parents, and other contacts;
2. characteristics – for collecting demographic data such as sex, race, and marital status;

²³ This software was produced by the Computer-Assisted Survey Methods Program (CSM) of the University of California at Berkeley, May 1998.

3. admissions – for collecting scores for undergraduate, graduate, and first-professional admissions tests;
4. enrollment – for collecting terms of enrollment, degree program, and field of study;
5. tuition – for collecting tuition data for the terms of enrollment;
6. financial aid awards – for collecting financial aid data for aid recipients;
7. need analysis – for collecting student financial aid budget data for aid applicants; and
8. ISIR – for collecting name and SSN for students not previously matched successfully to CPS, but for whom an ISIR was available, indicating the student had applied for federal financial aid for the study year.

Because the Web-CADE database was resident on an RTI Web server, daily status reports summarizing the progress of the Web-CADE institutions were generated and posted on the Integrated Management System (IMS). However, periodic calls were placed to the coordinators to inquire as to their progress, thereby prompting the institutions to complete the record abstraction. In general, status reports indicated that schools were typically slow in beginning the CADE task (often waiting many weeks after system initialization before starting data collection), but once they began they tended to complete the task relatively quickly.

Student CATI/CAPI interviews. Student interviews were conducted primarily by telephone, and occasionally in person, using CATI/CAPI technology. Like CADE, CATI/CAPI was developed using version 4.3 of the Computer-Assisted Survey Execution System (CASES) software to facilitate preloading full-screen data entry and editing of “matrix-type” questions. The CATI/CAPI system presented interviewers with screens of questions to be asked of the respondents, with the software guiding the interviewer and respondent through the interview, automatically skipping inapplicable questions based on prior response patterns or suggesting appropriate wording for probes should a respondent pause or seem uncertain in answering a question.

To reduce interview burden and to guide the interview through appropriate branchings (e.g., questions appropriate only for graduate students), considerable information was preloaded into the CATI records before the interviews. Such preloaded information included (1) data previously collected through CPS and/or CADE; and (2) information from the sampling file (e.g., name, Social Security number, school name, school and student stratum). In a number of instances, specific questionnaire items were not asked (or were only verified) if that information had been collected previously. Data were preloaded into CATI on a flow basis, as CADE results were received from the institutions.

Features of the CATI system that facilitated smooth and appropriate conduct of the interview included:

- extensive use of appropriate branching of interviewees based on preloaded information or responses to questions asked previously in the interview;

- extensive use of “fill” features in screen presentations of questions to be asked by interviewers (i.e., filling in part of a question with preloaded data or a previously provided response—that is, instead of asking the respondent something about “second postsecondary institution that they attended,” the question would be presented with the name of the institution embedded in the screen wording);
- a “breakoff/resume” feature allowing interview continuation after a breakoff to move automatically to the next applicable question for the respondent; and
- provision of context-sensitive “help” screens (available with a single keyboard entry) to provide the interviewer with information about particular questions to help clarify the question's intent.

Additionally, online coding programs developed by NCES (for industry/occupation, IPEDS, and field of study coding) were embedded in the overall interview administration system. These allowed standard coding of verbatim responses while the respondent was still available to assist.

The student CATI interview consisted of seven sections that were administered sequentially (see figure 2-3).²⁴ The sections were ordered so that important information was collected early in case the respondent broke off the interview before completion. A facsimile student interview is provided in appendix D.

Cases not completed in CATI (i.e., refusing and/or unlocatable cases) were assessed for assignment to field staff. If the case was in an identified geographic cluster, it was assigned to a field interviewer. The field interviewer then attempted to locate the student and complete the interview using CAPI. If the case was not in an identified cluster, it was assigned to a field locator. The field locator then attempted to locate the student and convince the student to call an 800 number to complete the interview in CATI.

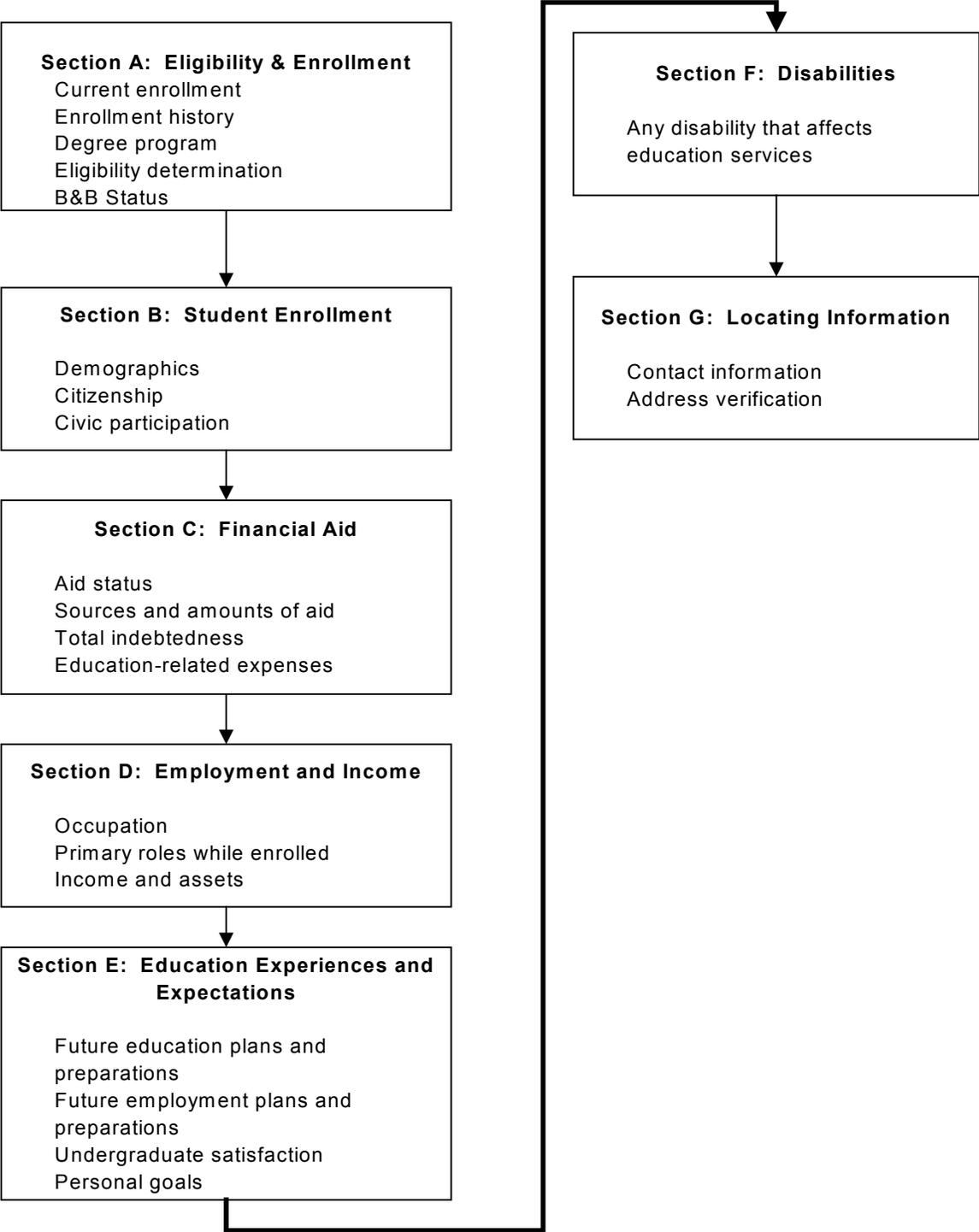
Results of CATI and CAPI interviewing were monitored daily through the study Integrated Management System (IMS). Daily reports of production, with revised projections of future production to satisfy study requirements, were available to both NCES and contractor staff.

Two sets of abbreviated interviews were conducted in special cases. First, the planned reliability reinterview study used an interview containing only a small subset of the items in the full student interview. Second, an abbreviated interview was developed in English and Spanish (containing only selected items) for telephone administration to those who were Spanish-speaking only²⁵ sample members or for use in refusal conversion. Facsimiles of the reliability interview and the abbreviated interview are provided in appendix F.

²⁴ While the logical flow within an interview is generally constrained to be linear (with forward branching as applicable), this is even more important in CATI, where previously supplied responses control subsequent branching items. Nonetheless, standard features were available to allow interviewers to back up in the interview to change prior responses based on information provided subsequently.

²⁵ Spanish speakers who could speak some English were guided through the full interview by bilingual interviewers. However, translation “on the fly” of the full interview to one who spoke only Spanish was considered inappropriate, and thus the Spanish translation of the abbreviated interview was administered in these cases.

Figure 2-3.—Structure and flow of NPSAS:2000 student CATI



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

Other *post hoc* student record data obtained. The electronic data interchange with the National Student Loan Data System (NSLDS), (including both loan and Pell grant files), ACT database, and ETS SAT files was initiated toward the end of CATI operations. As with the previously described procedures with CPS, matching of students to these files required Social Security numbers. At the time of these requests, apparently valid SSNs were available for 69,449 sample members, the number subsequently submitted for all attempted matches and associated data downloads.²⁶ In addition to SSNs, name and date of birth were submitted to ETS for SAT matching and to NSLDS for Loan and Pell matching. For ACT, sex and date first enrolled (if available) were included in the file along with name and date of birth. These variables assisted the data vendors in performing confirmatory data quality checks. All matching processes were initiated by RTI staff providing a file with one record per sample member with the requested data on a CD-ROM to the database system. A successful match with the NSLDS loan and Pell database required that the student have a valid application record within the database. Similarly, a successful match with the ACT and SAT databases required that the student have a valid record with the test databases. Additional data (e.g., date of birth) was used when necessary to increase the likelihood of a successful and accurate match.

2.3.2 Student Locating

The basic NPSAS:2000 design involved tracing sample members to their current location prior to conducting a computer-assisted telephone interview or a computer-assisted personal interview with them.

Pre-CATI locating. Locating information obtained during the institutional CADE phase of the study was incorporated into the locator database. The data files were updated in batch mode to the National Change of Address (NCOA)²⁷ system and Telematch²⁸ on a flow basis. After the locator database had been updated with the new information, a lead letter packet was mailed to the best address for the sample member. This mailing included a standard lead letter and a study leaflet. These mailings occurred on a flow basis twice a week beginning in May 2000 and continued throughout the data collection period. The most current information for the student and any other contacts were then preloaded into the CATI system to assist the interviewers in locating the sample members.

CATI-internal locating. When assigned a case, the telephone interviewer called the telephone number designated by the system as the best number (i.e., the number among all available locator numbers that appeared to have the greatest potential for contacting the sample member) and attempted to interview the designated sample member. When the person answering the call said that the sample member could not be reached at that number, the interviewer asked the person how to contact the sample member. If this query did not provide

²⁶ Of these, 8,120 were ultimately determined to be nonrespondents.

²⁷ NCOA is a database consisting of change of address data submitted to the U.S. Postal Service. Almost 100 million records are updated every 2 weeks and stored for 3 years.

²⁸ Telematch is a computerized residential telephone number look-up service consisting of over 65 million listings, over one million not-yet-published numbers of new movers, and over 10 million businesses. Telematch uses a name, street address, and ZIP code as search criteria and Reverse Telematch uses telephone numbers as the search criteria to provide the names under which telephones are listed.

the information needed, the interviewer initiated tracing procedures, using all information available to call other contact persons in an attempt to locate the sample member. When all tracing options available to the interviewer were exhausted without success, the case was assigned to intensive tracing via FastData²⁹, Tracing Operations Unit (TOPS)³⁰, or field interviewers/locators. The latter two intensive tracing steps are described below.

Intensive locating (post-CATI tracing). All cases that were not located during the CATI locating process were submitted to TOPS for intensive locating. TOPS implemented a two-tiered intensive tracing plan. The first tier involved identifying sample members with Social Security numbers and processing that information through a series of electronic databases. The specific tracing activities are listed below, and were restricted to the collection of locating/directory information.

- *Query of Equifax database.* Equifax is a credit bureau that maintains credit files on a large number of individuals.
- *Query of Internet databases.* Contractor staff had direct electronic access to various databases, which included names, Social Security numbers, and current and former addresses and telephone numbers of individuals.
- *Query of the Select Phone Book CD ROM data.* This database contains every published telephone number in the United States, with associated names and addresses. It can be sorted within city by address, to obtain telephone numbers and names of neighbors.

If the searches generated a new telephone number, that case was sent back to RTI's Telephone Survey Department (TSD) for telephone interviewing. If a new address was generated, but no telephone number, tracers called Directory Assistance or accessed other databases to obtain telephone numbers for the TSD. This first level of effort minimized the time that cases were out of production.

All remaining cases (those lacking new information from the SSN search) underwent a more intensive level of tracing in the second-tier approach. This approach involved the following procedures: (1) checking Directory Assistance for telephone listings at various addresses; (2) using electronic reverse-match databases to obtain the names and telephone numbers of neighbors and then calling the neighbors; (3) calling persons with the same unusual surname in small towns or rural areas to see if they were related to or knew the sample member; (4) contacting the current or last-known residential sources such as neighbors, landlords, current residents, tax assessors, realtors, and other business establishments related to previous addresses associated with the sample member; (5) calling colleges, military establishments, and

²⁹ FastData is a series of database searches used to locate sample members after pre-CATI batch database searches have been done but before sending cases for intensive interactive tracing.

³⁰ The Tracing Operations Unit (TOPS) is a highly specialized unit within RTI that was created in response to the recurring needs of certain research methodologies to locate large numbers of sample members. The sole focus of TOPS is tracing sample members so that they can be located for research studies; the unit does not involve any data collections.

correctional facilities to follow up on leads generated from other sources; and (6) checking various tracing Web sites. Tracers checked new leads produced by these tracing steps to confirm the address and telephone numbers for the sample members. When the information was confirmed, the case was returned to CATI for completion. If the information could not be confirmed (e.g., there were no working telephone numbers or numbers for relevant neighborhood sources were unpublished), the case was sent to the field.

Field locating. The main purpose of the intensive field locating/interviewing effort was to increase the response rate. However, since the costs of conducting these operations were high, field efforts were implemented only when less costly efforts were exhausted. Sample members were identified as needing field locating/interviewing if they were not located using CATI-locating and centralized intensive tracing.

Geographic clusters of sample members were designated, and 33 of these clusters were staffed with field interviewers who were trained to locate sample members and interview them using a laptop computer. Field cases falling outside the geographic clusters were assigned to field locators (trained as interviewers on other RTI studies) who located sample members in their local areas and encouraged them to call in to RTI's TSD to be interviewed.

2.3.3 Telephone Interviewing

CATI locating and interviewing began on May 22, 2000, and continued through February 28, 2001. CATI procedures included attempts to locate, gain cooperation from, and interview study sample members by telephone.

Before the CATI sequence began, notification letters on U.S. Department of Education stationery and with attachments were mailed to students. These letters notified the sample members of the upcoming survey, pointed out the importance of the study, disclosed average time burden, and urged participation.

Associated with the interviewing was the necessity (due to incomplete or incorrect telephone numbers), in many cases, to locate the respondent(s). Much of the locating challenge was associated with the fact that many NPSAS:2000 sample members (particularly those who had just received their degrees) were at a stage in their lives in which they were highly mobile. To facilitate the tracing component, each CATI record contained roster lines for up to 15 telephone numbers; each such roster line was associated with a history of the dates and results of all calls made to that number and a number-specific comment field. Up to five roster lines were preloaded with contact information. New roster lines were added during CATI tracing operations as a result of locating sample members via intensive tracing efforts. Locating calls were initiated according to a calling plan using an automatic call scheduler embedded within the CATI software. This system allowed calls to be scheduled on the basis of established case priority, time of day, and history of success of prior calls at different times and on different days.

Once located, an attempt was made to conduct the full interview with the sample member. However, some cases required special treatment. To deal with those who initially

refused to participate (including locator sources who acted as “gatekeepers,” preventing access to the sample member), certain interviewers were trained in refusal conversion techniques. Sample members and their locator sources who spoke only Spanish, primarily located in Puerto Rico, were assigned to bilingual CATI interviewers.

Finally, in an effort to increase study response rates, a modest incentive was used with particular types of nonrespondents: (1) cases where the sample member initially refused the interview; (2) sample members for whom intensive tracing yielded a good mailing address, but no telephone number; and (3) cases identified as “hard to reach” (i.e., those with 20 or more call attempts, where contact had been established with the sample member and no “hard” appointment was pending). The incentive consisted of a letter from the project director on RTI letterhead, tailored to the specific type of nonrespondent (i.e., refusal or hard to reach/no telephone number). A \$5 bill was included with the letter. Respondents were promised a check for \$15 if they called an 800 number to complete the interview. The incentive letters were mailed on a flow basis as respondents met one of the three criteria described above. All cases sent to field interviewers or field locators were automatically made eligible to receive the incentive once the case was sent to the field. Interviews were obtained from about half of the sample members who were offered the incentive with almost 60 percent of those initially refusing being converted by the incentive offer.

2.3.4 Field Interviewing

Field interviewing activities began after training was conducted and field cases and bulk supplies were shipped to the field interviewers. CAPI procedures included attempts to locate, gain cooperation from, and interview study sample members either by telephone or in person.

All students who were finalized in CATI and by TOPS as “unlocatable” were eligible for assignment to the field for CAPI interviewing or field locating. Sample members who had not completed the NPSAS:2000 interview at the time field interviewing began and who resided in an identified geographic cluster in the vicinity of a field interviewer were immediately assigned to the field. The field interviewer then attempted to locate the student and complete the interview using CAPI. If the case was not in an identified cluster, it was assigned to a field locator. The field locator then attempted to locate the student and convince the student to call an 800 telephone number to complete the interview in CATI.

Field interviewers documented every telephone call or field contact. They were provided with a checklist that included example questions to help with tracing operations and that demonstrated the correct order in which tracing activities should be performed. The checklist was completed for each case to help identify the sources that were most useful in locating the students.

Primary tracing sources included parents, current or former neighbors or roommates, the NPSAS school, and city and county offices. Secondary tracing sources included Directory Assistance, the Chamber of Commerce, public libraries, the U.S. Postal Service, and the Department of Motor Vehicles. Other miscellaneous sources for field interviewers, useful in some cases, included small town police or sheriff’s departments, fire departments or emergency rescue squads, local newspapers, public housing authorities, mobile home park managers, motel

staff, probation officers, and permit-issuing departments at the city level (new construction). A contact script guided interviewers in soliciting information from various sources.

When field interviewers successfully located sample members, they introduced themselves and explained the purpose of the study, referring to the advance letter mailed previously. They then attempted to complete the interview using the same instrument used in the CATI interview. The field staff were supported by a computerized control system that tracked field assignments and captured pending and final result codes. Daily reports, posted to the IMS, tracked the progress of the field effort.

2.3.5 Training CADE Data Collectors

The training for RTI CADE staff was held in two sessions to allow for efficient use of the field staff immediately following training. Prior to these sessions, six Field Supervisors hired for the CADE collection were trained in February 2000. The initial training for 23 CADE Field Data Collectors was conducted during April 2000. The second session was originally planned for June; however, this session was postponed to late July 2000 to coincide with the projections of list receipt from institutions, sample selection, and flow of cases into CADE. Staff scheduled to attend the June session were notified of the delay and there were no attrition problems related to the postponement. Five of the six Field Supervisors attended and participated in the training session and 13 Field Data Collectors successfully completed the session. To reduce travel costs for the relatively small number of trainees, the training sessions were held in the Research Triangle Park area.

The Field Supervisor training included a 2-day session on the background of NPSAS:2000 (including objectives, time frame, and the financial aid process), supervisory and administrative responsibilities, procedures for recruiting field data collectors, and use of the Case Management System, the assignment and transfer (WebATS) system, and the e-mail system. The Field Data Collector training included a half-day of training on the computer for a subset of the trainees (who needed an introduction to the computer) prior to the project training. Training consisted of an overview of the NPSAS:2000 objectives and time frame, explanation of how the financial aid process works on campuses, review of the architecture and nature of the CADE software, review of and practice with each section of the CADE instrument, procedures for contacting and dealing with the Institutional Coordinator and other staff at the institutions, instruction in and practice with locating records (including, but not restricted to use of the “location of records” lists provided by the Institutional Coordinators and review of ISIRs, procedures for contacting Field Supervisors, electronic transmission of completed cases, and administrative procedures.

During this training, considerable use was made of location and abstraction of records using mock student folders developed, with the assistance of NASFAA staff, to represent diversity in record keeping at different types of postsecondary institutions. Laptop computers were provided to all trainees for their use during training and subsequent field work. Additionally, as a training aid, each trainee was issued a *Field Data Collector Manual*³¹ and a

³¹ RTI *Field Data Collector Manual: NPSAS:2000 Main Study*. Research Triangle Park, NC, March 2000.

CADE Users' Guide.³² The tables of contents for both of these manuals as well as a copy of the Field Data Collector training agenda are included in appendix D.

Training of institutional staff in use of the Web-CADE application relied heavily on self-training, since the major objectives of that training were to become familiar with the CADE program and to learn how to access the program through the World Wide Web. A secure user ID and password were required to access the system. Help screens were embedded within the program and a “hotline” number and e-mail address were established through which users could obtain answers to specific or general questions from RTI central office staff who developed the software. Additionally, institutional staff were provided with a copy of the *CADE Users' Guide*.

2.3.6 Training of CATI/CAPI Interviewers and Tracing Specialists

The mixed-mode design of the NPSAS:2000 student data collection required the development of three separate training programs: CATI interviewing, field interviewing, and tracing. Each training program consisted of separate protocols for data collectors and for supervisors. For each, training topics covered administrative procedures, including confidentiality requirements and quality control techniques; student locating; interactions with students, parents, and other contacts; the nature of the data to be collected; and the organization and operation of the CATI, CAPI, and tracing operations systems used for data collection. The goals for these training programs were to

- increase the accuracy, quality, and timeliness of the data collected;
- standardize the quality of data collection techniques and procedures; and
- provide explicit, nonjudgmental procedures for telephone interviewers, telephone monitors, field staff, tracing specialists, and supervisors to follow.

Training telephone interviewers. Initial training for telephone interviewers, monitors, and supervisors began in late April 2000 immediately before student data collection started. Most of the supervisors and monitors used on the project were trained in a separate session, prior to interviewer training, so that they could assist during subsequent training sessions. Because cases flowed into CATI over time from the school data collection effort (rather than being loaded all at once at the outset of data collection), it was necessary to schedule the required training sessions over time to mirror the CATI workload. In all, 23 training sessions were held for CATI interviewers, monitors, and supervisors between April and December 2000. In total, 372 telephone interviewers were trained over this 9-month period. Table 2-6 lists the training sessions offered and the number of interviewers, supervisors, and monitors completing each training program.

³² RTI *CADE Users' Guide: NPSAS:2000 Main Study*. Research Triangle Park, NC, April 2000.

Table 2-6.—CATI training activities and number of interviewers trained

Training activity	Number of sessions	Number of people trained
CATI supervisor/team leader training	1	37
CATI monitor training	1	20
General telephone interviewer training	16	297
Telephone interviewer training and refusal avoidance	20	372
Telephone interviewer refusal conversion training	8	86
Tracing specialist training	9	106

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

Newly hired interviewers with no prior telephone interviewing experience were also provided with 8 hours of general or introductory CATI training before they were allowed to attend the project specific training. In these sessions, new interviewers were instructed on general interviewing techniques and best practices, the screen layout and coding conventions used on all CATI projects conducted at RTI, and the routine administrative procedures and requirements for working in RTI’s Telephone Survey Department. New interviewers who did not successfully complete the 8 hours of general training were not allowed to proceed to the project-specific NPSAS:2000 training.

Project-specific training for CATI-experienced telephone interviewers and new hires who successfully completed general interviewer training consisted of 20 hours of classroom and practical, hands-on training. Topics covered included the nature and purpose of NPSAS:2000 and the B&B:2000/2001 follow-up; the procedures and protocols to be used for tracing, contacting, and interviewing sample members; and an extensive review of the NPSAS:2000 instrument. During the training, all questions in the interview were reviewed, and interviewers received both written and hands-on practice with the screens and subroutines for conducting online coding, and time for both group and individual practice with the instrument itself. Prescribed or “mock” interviews were designed to ensure that interviewers received hands-on practice with the most common paths through the questionnaire as well as practice administering some of the more difficult items in the questionnaire. Small group training, using audiotaped scenarios, was also provided to enhance refusal avoidance skills. At the end of the project-specific training, all interviewers were required to complete a certification process to ensure their readiness to conduct efficient and reliable interviews for the project. The certification process involved the successful administration of the NPSAS interview in a paired “mock” situation with a fellow trainee (one playing the interviewer, the other the sample member). Trainers monitored these sessions, noting any difficulties a trainee might have had with questionnaire administration; use of online coding programs; keying accuracy; and voice tone, speed, and quality. Those who did not successfully complete the training and pass the certification process were not allowed to work on the study.

At the outset of the training, each interviewer received a detailed NPSAS:2000 *Telephone Interviewer Manual*³³ that served as both an instruction guide for the training’s

³³ RTI *Telephone Interviewer Manual for NPSAS:2000*. Research Triangle Park, NC: April 2000.

lectures, discussions, and practical exercises, and as a reference guide for use after completion of training. The manual's table of contents and a sample of the training agenda for telephone interviewer training are included in appendix D. The interviewer manual, supplemented with additional materials more directly related to supervisory activities, was also provided to telephone supervisors and monitors.³⁴ The supplementary materials included data collection schedules and staff contact information, procedures for supervising interviewers during data collection, tracing review and other quality control activities, problem resolution, refusal avoidance and conversion techniques, and administrative and record-keeping activities.

Staff involved with interviewer monitoring received 2 hours of additional instruction on the protocols and procedures for conducting interviewer performance monitoring and quality assurance monitoring. The training included a review of the interviewer performance monitoring form and hands-on practice with the online program developed for quality assurance monitoring. Each monitor received a separate manual documenting the procedures to be followed.³⁵

Six weeks after the start of student interviewing, project staff began conducting a series of refusal conversion trainings for a subset of the highest-performing telephone interviewers. CATI supervisors and monitors evaluated the effectiveness of telephone interviewers in dealing with respondent objections and overcoming barriers to participation. The most effective interviewers received additional and specialized instruction in specific refusal conversion techniques, including obtaining cooperation from sample members, addressing concerns raised by parents and other sample gatekeepers, validating the importance of the study, and encouraging participation among sample members who were nonrespondents in the previous data collection. During the course of data collection, 86 interviewers completed refusal conversion training.

Training field interviewers. To ensure standardization and reliability in the field data collection effort, all field interviewing and supervisory staff were required to complete a 32-hour comprehensive training program designed to maximize both data quality and interview response rates. This training program included classroom lectures, hands-on practice, and other practical exercises. The content of the training sessions focused on an overview of the nature and purpose of NPSAS:2000 and the B&B:2000/2001 follow-up, procedures for tracing and contacting sample members in the field, an extensive question-by-question review of the NPSAS:2000 instrument, practice with the interview screens and online coding programs, and time for both group and individual practice.

As with the telephone interviewer training, the field interviewer training program provided hands-on training with the CAPI interview program. Additionally, the training program covered tracing techniques, contacting protocols, and case management, including the use of electronic mail and data transmissions systems, troubleshooting guidelines for the laptop computer, and field-specific reporting and administrative requirements.

³⁴ RTI *Telephone Supervisor's Manual for NPSAS:2000*. Research Triangle Park, NC: April 2000.

³⁵ RTI *Monitor Manual for NPSAS:2000*. Research Triangle Park, NC: April 2000.

Each interviewer received a copy of the NPSAS:2000 *Field Interviewer Manual*³⁶ at the start of the training. This manual, which served as both an instructional resource and a reference book for the field work, introduced and reviewed many topics important to the study. The classroom instruction, discussion, and practical exercises focused on general interviewing, field tracing, and student contacting. The manual and field interviewer training also provided instruction for reviewing the case history documentation generated by in-house tracing activities to avoid repeating steps taken during earlier tracing efforts (e.g., telephone interviewer contacts and centralized tracing efforts). The interviewer manual, supplemented with additional materials more directly related to supervisory activities, was provided to field supervisors.³⁷ The supplementary materials included data collection schedules and staff contact information, procedures for supervising interviewers during data collection, tracing review and other quality control activities, problem resolution, interview verification procedures, and administrative and record-keeping activities.

Initial training for field supervisors took place in August 2000, several weeks before the first field interviewer training session. These supervisors then assisted with the initial training for field interviewers that took place in September, just before the start of field data collection. Two more training sessions were held for additional field interviewers in November and December. Overall, 6 field supervisors and 74 interviewers completed the field interviewer training for NPSAS:2000.

Finally, 65 field locators, who were used to assist with tracing of unclustered nonrespondent cases, were trained using a home-study packet, rather than a centralized training program. As case assignments were made, each field locator was sent home-study materials consisting of a study overview, a field locator manual that explained the nature of the assignment and the steps to be followed in locating hard-to-find sample members, instructions for making contact with sample members and other potential contacts, and a set of example tracing materials. Field locator assignments were made initially in October 2000 and continued through January 2001.

Training tracing specialists. Staff working in RTI's TOPS on the centralized locating and tracing activities for NPSAS:2000 also received project-specific instruction, although not as extensive as the programs developed for telephone and field interviewers. Each tracing specialist received two hours of instruction, including an overview of the nature and purpose of NPSAS:2000 and the B&B:2000/2001 follow-up; the study schedule; protocols for contacting sample members, gatekeepers, and other contacts; the tracing steps and techniques to be used for locating NPSAS:2000 sample members; and the tracing-specific reporting and administrative requirements for the study.

Newly hired tracing specialists also received 8 hours of general tracing instruction. This training focused on general tracing techniques; use of the computer search resources in TOPS; documentation of locating steps in the TOPS case management system; techniques for obtaining locating information for sample members from parents, gatekeepers, and other contacts; and the

³⁶ RTI *Field Interviewer Manual for NPSAS:2000*. Research Triangle Park, NC, May 2000.

³⁷ RTI *Field Supervisor Manual for NPSAS:2000*. RTP, NC, May 2000.

general and routine procedures for working in the TOPS unit. Tracers who did not successfully complete the general tracing training were not permitted to attend the project-specific training.

Eight training sessions were held between May and November 2000 for tracing staff. In total, 8 tracing supervisors and 83 tracing specialists were trained to work on NPSAS:2000.

2.3.7 Evaluation and Quality Control Design

Each major component of the NPSAS:2000 full-scale study was evaluated. Formative evaluations were designed to assess tasks at intermediate stages so that the effects of employing alternate methodologies could be analyzed, and modifications and revisions could be employed and assessed prior to task completion. Other evaluations assessed the ultimate outcomes of the survey. A summary of NPSAS:2000 evaluations that were planned and implemented is provided in table 2-7.

As indicated in table 2-7, the study design included a number of components for evaluation of data quality. Among these, a reliability reinterview was conducted with students about 8-12 weeks after the initial interview; this involved a random subsample of respondents to the initial interview. The reliability reinterview contained only a small subset of the initial interview items. Also critical to the operational evaluation and quality control were the regular quality circle meetings with field interviewers, telephone interviewers, interview monitors, and interviewer supervisors. These meetings provided an easily available forum for production staff and project management to address the important topic of work quality, discuss issues of concern, identify problems with the survey instruments, share ideas for improving the instruments, and suggest various approaches for improving operations and/or results. To implement suggested improvements arising from these meetings, the operational features of the CATI instrument were sometimes refined over the course of the data collection period. On completion of data collection, final quality circle meetings were held, serving as debriefing sessions for the full operational period.

2.4 The Integrated Management System

The NPSAS:2000 IMS was developed based on a framework initially developed (and refined) under previous NCEs studies conducted by RTI. These include BPS:90/92, BPS:90/94, NPSAS:96, and BPS:96/98. As with these previous studies, the NPSAS:2000 IMS consisted of independent, but integrated, modules. Development of the IMS occurred throughout the study field test period, and was modified before the full-scale study based on field test results. To the extent possible, the NPSAS:2000 IMS was developed using commercial, off-the-shelf PC-based software systems.

Table 2-7.—Summary of NPSAS:2000 evaluations

Major area of evaluation	Evaluation approaches
Training	Debrief field abstractors.* Debrief CATI staff.*
Enrollment file acquisition	Analyze overall response rate, accuracy, costs, and time to produce lists.
Record abstraction	Evaluate electronic file matching/downloading approaches. Analyze data quality (missing data) under conditions of web-CADE, field-CADE, and data file production approaches. Debrief institutional coordinators.* Debrief field staff.*
Tracing activities	Debrief tracing staff and supervisors.* Analyze all levels of tracing results and costs.
Interview administration/data quality	Analyze silent monitoring quality control data. Analyze CATI operational parameters (e.g., numbers of calls per case, total interviewer hours per completed interview). Analyze interview response burden, overall and by section. Debrief interviewers, monitors, and supervisors.* Analyze response rates and patterns of interview nonresponse. Analyze impact of financial incentive on response rate. Analyze response temporal stability (reliability) through reinterviews of selected items. Analyze effectiveness of various strategies for handling answering machines.

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

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

The major enhancement to the NPSAS:2000 IMS was the development of a Web-CADE module for institutions to provide student data via the Internet. The system replaced the diskette-based version of CADE used during NPSAS:96. The Web-CADE system included encrypted data transmission and a login/logout feature to maintain data security. More information about Web-CADE is provided below.

The modular design of the IMS allowed for efficient upgrading or replacement of components, or modules, as necessary. This occurred during the field test period, as RTI’s migration from SQL Server 6.5 to SQL Server 7.0 took place during the summer of 1999.

Below are listed the major modules of the NPSAS:2000 IMS. Relevant details regarding each module are provided.

Receipt Control System (RCS)

- Back-end database is Microsoft SQL Server. SQL Server version 6.5 was used for the field test development. The RCS back-end database was upgraded to SQL Server version 7.0 near the end of the field test period and before the full-scale study.
- Front-end interface was programmed in Microsoft Visual Basic 5.0 and Microsoft Access 97.
- RCS reports were developed using Crystal Reports 6.0 and Microsoft Access 97.

Web CADE

- Back-end database was Microsoft SQL Server 6.5 (subsequently upgraded to version 7.0).
- Front-end interface was programmed in HTML.
- Middleware software, which allows the Web pages to communicate with the back-end database, was Allaire Cold Fusion version 4.0.
- Web-CADE edit checks were programmed using JavaScript.
- Reports were developed using Crystal Reports 6.0, Microsoft Access 97, and Cold Fusion 4.0.
- Web security was implemented using Secure Socket Layer (SSL) certification with 128-bit encryption. User IDs and passwords were assigned by RTI using Microsoft Windows NT 4.0 domain security.
- Users' browsers were required to support, and be enabled for, JavaScript.

Field CADE

- The field CADE system was run on Toshiba Satellite laptop computers configured with 32MB of RAM and Pentium processors.
- Back-end database was CASES version 4.3.
- Instrument was programmed in CASES 4.3.
- User exits were programmed using C++.
- Final CADE database was maintained in SAS version 6.12.
- CADE quality control reports and status reports were programmed in SAS 6.12.

CATI/CAPI

- Back-end database was CASES version 4.3.
- Main instrument was programmed in CASES 4.3.
- Abbreviated instrument (for use in refusal conversion and hardcopy format) was programmed in CASES 4.3.
- CATI user exits were programmed using C++.
- Final CATI database was maintained in SAS 6.12 (subsequently upgraded to SAS 8.1).
- CATI status and summary reports were programmed in SAS 6.12 (subsequently upgraded to SAS 8.1).
- The CATI system was ported to a CAPI version, for use in conducting in-person interviews with students. The same software systems were used for the CAPI system, with the exception of a case management component developed in SQL Server 7.0 and Visual Basic 5.0.

Data Library

- CD-ROM-based searchable database of Data Library entries was maintained in SQL Server 7.0 throughout the course of the study. The Data Library was initialized during the NPSAS:2000 field test.
- Web-based searchable database of Data Library entries was programmed in Cold Fusion 4.0 and Microsoft Access 97.
- Word processed documents were created using Microsoft Word.
- Spreadsheets were created in Microsoft Excel.
- Schedule files were maintained in Microsoft Project 98.

IMS Web site

- Infrastructure was programmed in HTML, with Cold Fusion 4.0 providing “action pages.”
- SQL Server 7.0 served as the back-end database where applicable (maintaining the project staff contact list, Technical Review Panel membership, confidentiality report, etc.)

Central Processing System (CPS)

- Back-end database for CPS data received was SAS version 6.12 and version 8.1.
- The CPS was a mainframe-based system called the Title IV Wide Area Network (T4WAN). Communications with T4WAN were through EDConnect for Windows version 2.3.
- CPS input files were prepared using SAS 6.12 / 8.1. Input files were flat ASCII files, with the Federal Data Request (FDR)-file layout (as specified in the CPS Electronic Data Exchange Technical Reference manual).
- CPS data files were read using SAS 6.12 / 8.1. CPS data files were flat ASCII files (one record per student, plus header and trailer records) with FDR full ISIR layout (as specified in the CPS Electronic Data Exchange Technical Reference manual).

National Student Loan Data System (NSLDS) processing

- Input files for matching to the NSLDS were created as flat ASCII files, containing student name, SSN, and date of birth. Files contained one record per sample student.
- NSLDS data were received as ASCII files containing loan-level transactions (multiple records per student). NSLDS loan records reflected cumulative history of loan data (i.e., not just the NPSAS year).
- Pell Grant data files were also received from NSLDS as flat ASCII files containing Pell-award-level records. As with the above-mentioned loan data, each student’s cumulative Pell history was obtained.
- All NSLDS input files were created and processed using SAS 6.12 / 8.1.
- Back-end database for all NSLDS data was SAS 8.1 format.

Admissions test file processing

- Student SAT data (scores and background variables) were obtained from ETS. ACT scores and background variables were obtained from ACT.
- Input files for submission to ETS and ACT were flat ASCII files, containing student name, SSN, and date of birth. Files contained one record per sample student.
- Admissions test files (received back from ETS and ACT) were flat ASCII files containing student-level records (one record per student). A separate file was received for each admissions test cohort year (multiple files received from each admissions test vendor).
- Input files for admissions test data were created and processed using SAS 6.12 / 8.1.
- Back-end database for admissions test data was SAS 6.12.

Automated processing

During full-scale data collection, a series of automated batch files were executed nightly via Windows NT scheduled processing. These automated processes included the following.

- Zero record update

Each night a process would run to copy the CATI “Zero” record (i.e., the master case status file) to an SQL table within the RCS database. This information was used to synchronize files between the RTI call center and the data being collected by field data collectors. The two key synchronization fields were the current status (interview complete, pending, refusal, etc.), and incentive group assignment (used to trigger incentive mailouts to “unable to locate” and “refusal” cases).

- Institution comments

This automated process updated the IMS Web site with searchable case-level comments from institution contacting staff. This provided the project team members with up-to-date information for use in communicating with institution staff.

- Master CADE upload

Each night this process would move CADE data from the public web CADE database to the master CADE database inside the RTI firewall.

- Dataload

This program contained many different subprocesses, with the overall purpose being to process transactions generated during the day by various project systems and activities, and post the transactions to the Receipt Control System, updating

institution and student-level case status information. Transactions included results from enrollment list processing, sampling, CPS matching, CADE preload and data receipt processing, lead-letter mailout and return, and CATI/CAPI preloading and interviewing.

- RCS report generator

Each night following the completion of the dataload process, the RCS report generator created HTML pages detailing both the institution- and student-level current status reports. It also produced miscellaneous project management reports including: Abstraction Method Report, Enrollment (list type) Report, Chief Administrator Participation Report, Enrollment List Acquisition Report, CADE Status Summary Report (overall and for the B&B cohort), and CATI/CAPI Summary Reports. The process automatically posted these reports to the IMS.

