

# Chapter 1

## Introduction, Background, and Purpose

---

This document describes, summarizes, and evaluates the methodological procedures and results for the full-scale 2000/01 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01). B&B:2000/01 was the follow-up study of bachelor's degree recipients identified during data collection in 2000. Research Triangle Institute (RTI), assisted by MPR Associates, Inc., conducted the study for the National Center for Education Statistics (NCES) of the U.S. Department of Education, as authorized by Title IV, Section 401 of the National Education Statistics Act of 1994 (P.L. 103-382).

This introductory chapter briefly describes the background, purposes, field test, and the schedule of activities and products of B&B:2000/01. In chapter 2, the study design and method are described. Overall outcomes of data collection are presented in chapter 3. Chapter 4 documents the quality of the data collected. Chapter 5 explains the methods used for variable construction and development of the data files. Chapter 6 provides a summary of the weighting procedures and estimate precision for the study. Materials used during the study are provided as appendixes to the report and cited, where appropriate, in the text.

### 1.1 Background and Purpose of the Baccalaureate and Beyond Study

B&B was designed to provide information of interest regarding higher education in the United States to policymakers, educators, and researchers. The focus of the study is on all types of baccalaureate degree holders, providing information regarding the supply of people trained in a variety of fields including education, science, and technology. As in the previous B&B study, base year data for potential B&B respondents were collected during National Postsecondary Student Aid Study (NPSAS) data collection.

NPSAS:2000, the base year collection for B&B:2000/01, was a comprehensive, nationwide study designed to determine how students and their families pay for postsecondary education and to provide demographic information on the postsecondary student population in the United States. NPSAS has been conducted since 1987 at 3- and 4-year intervals. NPSAS:93 served as the base year of data collection for the first Baccalaureate and Beyond Longitudinal Study (B&B:93/94). Students who were identified in NPSAS:93 as baccalaureate recipients during the 1992–93 school year were interviewed again in 1994 and 1997, and will be interviewed a fourth time in 2003. However, there will be no further follow-up with the B&B:2000/01 cohort.

One of the major issues addressed by the B&B studies is the length of time taken by undergraduates to earn bachelors' degrees. B&B:2000/01 permits continuation of the "time to degree" trend data provided by the preceding NCES Recent College Graduates (RCG) cross-

sectional study series and the B&B:93/94 longitudinal surveys. These data are periodically reported as indicators in the annual *Condition of Education* published by NCES.

Two other pertinent study issues concern the economic value of obtaining a bachelor's degree and the debt associated with degree attainment. B&B:2000/01 extends trend data from RCG and B&B:93/94 on employment status and income 1 year after graduation. The issue of student debt, through both federal student loan programs and the increased use of credit cards, is particularly relevant for recent college graduates. Data collected for B&B:2000/01 provide an estimate of the average debt accumulated by students in order to complete a bachelor's degree, and how this may affect students' future plans. With the addition of federal student loan history data from the U.S. Department of Education's National Student Loan Data System (NSLDS), B&B:2000/01 expands and improves information on student debt and repayment.

B&B:2000/01 gathered a variety of data on post-baccalaureate education, such as the demographics of those attending, fields of study, and whether these students also participated in the work force. It also offers an important opportunity to study early outcomes of newly qualified teachers, including entry and attrition, certification, and participation in the teaching profession, thus enabling the continuation of a series of earlier NCES teaching reports. B&B:93/94 provided the data for the NCES report on newly qualified teachers.<sup>1</sup> B&B:2000/01 will again collect these data and offers the opportunity for comparative analyses. These data can be used to determine whether recent graduates' perspectives regarding teaching or the likelihood to enter teaching have changed in the 7 years between cohorts. In addition, B&B:2000/01 provides information about the supply of graduates trained in science and technology, and about the graduates who are employed in those fields.

## **1.2 Overview of B&B:2000/01**

B&B:2000/01 occurred in two parts: (1) a field test and (2) the subsequent full-scale study, both of which are described below. There were two eligibility requirements for participation in the B&B:2000/01 full-scale study. First, a person had to have been enrolled at the sampled NPSAS school at some point during the 1999–2000 academic year; and second, either the student or institution must have indicated receipt or expected completion of a baccalaureate degree within the designated time frame (July 1, 1999, to June 30, 2000). Survey data were collected using computer-assisted telephone interviews (CATI). Over 10,000 respondents completed interviews for this study, with data collection occurring between July 9, 2001, and November 20, 2001.

---

<sup>1</sup> Henke, R.R., Geis, S., Giambattista, J. (1996). *Out of the Lecture Hall and Into the Classroom: 1992-1993 College Graduates and Elementary/Secondary School Teaching, With an Essay on Undergraduate Academic Experiences* (NCES 96-899). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

### 1.3 Summary of B&B:2000/01 Field Test

The field test data collection period for the B&B:2000/01 was March 2000 to July 2000. The major purpose of the field test was to evaluate all operational and methodological procedures, instruments, and systems proposed for use in the first follow-up study. The field test sample for B&B:2000/01 comprised 860 students from the NPSAS:2000 field test sample. Data were collected from 700 respondents, resulting in an 81 percent response rate. Based on the results of this field test, several changes were made to the B&B:2000/01 full-scale study. These included the following:

- a change in the eligibility requirements for participation,
- the deletion of selected survey items,
- the modification of survey items based on field-test responses, and
- a reduction in the use of nonresponse incentive payments.

Previous cycles of NPSAS had required baccalaureate degree recipients to have been enrolled at the NPSAS institution during the NPSAS year. In the B&B:2000/01 field test, all baccalaureate degree recipients during the NPSAS year were considered potential sample members, regardless of whether they were enrolled at the NPSAS school during the NPSAS year. However, during the field test, it was determined that some questions were difficult to answer for those respondents who had received their degree at a school other than the NPSAS school. Therefore, the change in eligibility implemented in the B&B:2000/01 field test was reversed, effectively keeping eligibility requirements the same for the B&B:2000/01 full-scale study as has been used in the prior B&B studies.

Several items were deleted from the full-scale study because they were deemed unnecessary for analysis. Typically, item revisions involved adding more response categories (see appendix B for revised data elements). Other item modifications included changing item wording so it was more specific or easier for respondents to understand. For a complete summary of the B&B:2000/01 field test study and an in-depth description of the recommended changes to the current full-scale study, readers are encouraged to refer to the field test methodology report.<sup>2</sup> It is available online at the National Center of Education Statistics web site (<http://www.nces.ed.gov/pubsearch/pubsinfo.asp?pubid=200115>).

---

<sup>2</sup>Biber, M.R., Link, M.W., Riccobono, J.A., and Siegel, P.H. (2001). *Baccalaureate and Beyond Longitudinal Study: 2000/01, Follow-up Field Test Methodology Report* (NCES 2001-15). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

#### **1.4 Schedule and Products for B&B:2000/01**

B&B:2000/01 full-scale data collection took place between July 9 and November 20, 2001. Table 1.1 presents the operational schedule for the B&B: 2000/01 follow-up study.

**Table 1.1. Start and end dates for major B&B:2000/01 study activities**

<b>Activity</b>	<b>Start date</b>	<b>End date</b>
NPSAS:2000 data collection	May 22, 2000	February 28, 2001
B&B:2000/01 field test data collection	March 28, 2000	July 2, 2000
B&B:2000/01 Follow-up study		
-Pre-load base year data into CATI records	June 21, 2001	June 21, 2001
-CATI training	July 5, 2001	July 8, 2001
-Mailed student prenotification letters	June 27, 2001	June 28, 2001
-CATI data collection	July 9, 2001	November 20, 2001

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

A number of reports and data products are anticipated for the B&B:2000/01 study. These include the following:

- Electronically documented, restricted access research files (with an associated electronic codebook [ECB]) for research data users. The ECB is a tool that gives the user a vehicle to browse through the lists of variables and variables’ descriptions for data sets. It also produces SAS or SPSS programs to access the data, Access files, and printable codebooks.
- A Data Analysis System (DAS) for public access to B&B:2000/01 data. The DAS is a Windows software application that provides public access to survey data files. The DAS allows users to specify the information they would like to have displayed in tabular form and, to some degree, the format of the resulting tables (which are fully labeled and usable). It can also create correlation matrices that can be used for higher-level statistical analyses.
- This methodology report for the follow-up study, providing details of sample design and selection procedures, data collection procedures, weighting methodologies, estimation procedures and design effects, and the results of nonresponse bias analyses.
- Two descriptive summaries: (1) a B&B descriptive summary with an overview that will focus on time to degree, labor market experiences, entry into graduate school, and household demographics; and (2) a report on newly qualified teachers that will compare the undergraduate experiences of new teachers with those of non-teachers.

# Chapter 2

## Design and Method of the Follow-up Study

---

A summary of the procedures used for the design and implementation of B&B:2000/01 is provided in this chapter. A brief explanation of the sampling methods used in NPSAS:2000 is included, along with detailed discussions of the subsequent sampling for B&B.

### 2.1 Sample Design and Selection

The B&B:2000/01 sample design had three stages, as illustrated by figure 2.1. The first two stages occurred within the NPSAS:2000 sample: a sample of NPSAS-eligible institutions and a sample of students within institutions. The third stage was B&B:2000/01 specific and provided subsamples from confirmed and potential baccalaureate recipients identified during the second stage of the NPSAS:2000 sample. In this chapter, an overview of each of these three stages is provided. Only a brief overview of the first two stages is presented because they are part of NPSAS:2000, which is described in detail elsewhere.<sup>3</sup> The third stage, unique to B&B:2000/01, is described in detail here.

#### 2.1.1 Target Population

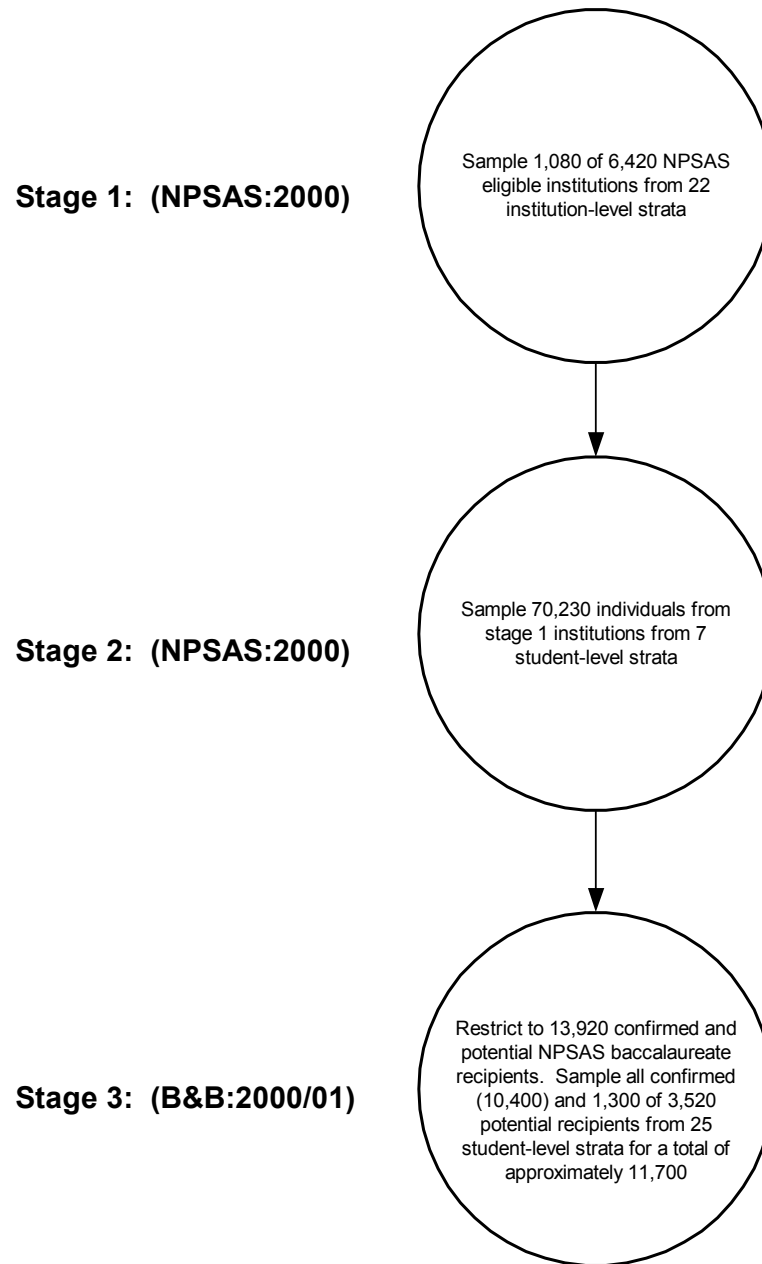
To define the target population for the B&B study, both eligible institutions and eligible students within these institutions need to be defined. Eligible institutions are those that satisfied the NPSAS eligibility criteria in the 1999–2000 academic year. An institution satisfied these criteria if the institution:

- offered an educational program designed for individuals who had completed secondary education;
- offered more than just correspondence courses;
- offered at least one academic, occupational, or vocational program of study lasting at least 3 months or 300 hours;
- offered courses open to more individuals than the employees or members of the company or group (e.g., union) that administered the institution;
- was located in the 50 states, the District of Columbia, or Puerto Rico;
- was not a U.S. service academy;

---

<sup>3</sup>Riccobono, J. A., Cominole, M.B., Siegel, P.H., Gabel, T.J., Link, M.W. and Berkner, L.K. (2001). *National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), Methodology Report* (NCES 2002-152). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Hereafter referred to as Riccobono et al. 2001.

**Figure 2.1. Stages of the B&B 2000/01 sample**



NOTE: Numbers have been rounded in accordance with NCES statistical standards in order to protect the confidentiality of respondents.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

- was open to the public; and
- was a U.S. Department of Education Title IV participating institution.<sup>4</sup>

Eligible students were individuals who were enrolled at, and obtained or expected to obtain a baccalaureate degree from an eligible institution between July 1, 1999, and June 30, 2000. Numbers of students and institutions have been rounded in this report in accordance with NCES statistical standards to protect the confidentiality of respondents.

### **2.1.2 The First Stage of the NPSAS:2000 Sample**

The first stage of the NPSAS:2000 sample was a stratified sample of institutions, with a sampling frame derived from eligible institutions contained within the 1998–1999 Integrated Postsecondary Education Data Systems (IPEDS)<sup>5</sup> of all Title IV participating institutions of the United States and its territories. The variables used to define the strata were institutional control, highest level of offering, and percentage of baccalaureate degrees awarded in education. The crossing of the levels of these variables produced 22 non-empty strata from which to sample. Sample sizes within each stratum were developed in consultation with NCES.

Given the agreed-upon allocation, institutions were sequentially sampled within each stratum with probabilities proportional to size (pps) using a probability minimum replacement (pmr) algorithm (Chromy, 1979).<sup>6</sup> Summary statistics of the final institutional sample are given in table 2.1. From this table, it can be seen that about 1,080 of the total of approximately 6,420 institutions were selected into the sample. Of the selected institutions, about 290 with the largest eligible student bodies were certainty institutions, and so were included with probability equal to one. Further breakdown of the institutional sample according to stratum is provided in the table.

### **2.1.3 The Second Stage of the NPSAS:2000 Sample**

The second stage of the NPSAS:2000 sample was a stratified systematic sample of individuals within sampled institutions. There were seven student strata: baccalaureate business, baccalaureate non-business, other undergraduate, masters, doctoral, other graduate, and first-professional. The information needed to identify students within these strata was provided by the sampled institutions.

Target stratum sample sizes were found by first obtaining population totals for the 7 student strata crossed with the 22 institutional strata from the IPEDS census. These population totals allowed for the development of a proportional sample allocation, in which the allocations were further refined after discussion with NCES. Within strata for a given institution, students

---

<sup>4</sup> Institutions participating in Title IV programs are accredited by an agency or organization recognized by the U.S. Department of Education, have a program of over 300 clock hours or 8 credit hours, have been in business for at least 2 years, and have a signed Program Participation Agreement (PPA) with the Office of Postsecondary Education (OPE), U.S. Department of Education.

<sup>5</sup> IPEDS is described as "a comprehensive system that collects institutional data about all primary providers of postsecondary education, and is built around a series of interrelated surveys designed to collect institution-level data in such areas as enrollments, program completions, faculty, staff, and finances." See the NCES Handbook of Survey Methods, NCES 2003-603 (p. 138).

<sup>6</sup>Chromy, J.R. (1979). Sequential sampling methods. *Proceedings of the American Statistical Association Section on Survey Research Methods*, 401–406. Hereafter referred to as Chromy 1979.

**Table 2.1. Institutional sampling rates and number of certainty and noncertainty institutions sampled, by institutional stratum for NPSAS:2000**

Institutional stratum <sup>1</sup>	Size of universe <sup>2</sup>	Sampling rate	Number of sample institutions		
			Total <sup>3</sup>	Certainty	Noncertainty
<b>All institutions</b>	6,420	0.17	1,080	290	800
<b>Public</b>					
1 Less-than-2-year	260	0.14	30	10	30
2 2-year	1,210	0.16	200	10	190
<b>Total less-than-4-year</b>	1,460	0.16	230	20	220
3 Bachelor's high education	20	0.29	10	0	10
4 Bachelor's low education	70	0.27	20	0	20
5 Master's high education	50	0.49	30	0	20
6 Master's low education	200	0.40	80	10	70
<b>Total 4-year non-doctorate-granting</b>	330	0.38	130	10	120
7 Doctorate-granting high education	30	1.00	30	30	0
8 Doctorate-granting low education	80	0.77	60	30	30
9 First-professional-granting high education	30	1.00	30	30	0
10 First-professional-granting low education	120	0.89	100	90	20
<b>Total 4-year doctorate-granting</b>	250	0.88	220	170	50
<b>Private not-for-profit</b>					
11 Less-than-2-year	110	0.10	10	0	10
12 2-year	310	0.07	20	#	20
<b>Total less-than-4-year</b>	420	0.08	40	#	30
13 Bachelor's high education	110	0.15	20	0	20
14 Bachelor's low education	400	0.09	40	0	40
15 Master's high education	120	0.31	40	0	40
16 Master's low education	410	0.20	80	10	80
<b>Total 4-year non-doctorate-granting</b>	1,050	0.16	170	10	170
17 Doctorate-granting high education	20	0.66	20	10	10
18 Doctorate-granting low education	90	0.31	30	#	20
19 First-professional-granting high education	80	0.71	60	30	30
20 First-professional-granting low education	290	0.23	70	30	30
<b>Total 4-year doctorate-granting</b>	490	0.35	170	80	90
<b>Private for-profit</b>					
21 Less-than-2-year	1,390	0.06	80	0	80
22 2-year or more	1,030	0.05	50	#	50
<b>Total private for-profit</b>	2,410	0.05	130	#	130

# Rounds to zero. All zeros provided in this table are actual values.

<sup>1</sup>Stratum reflects institutional categorization as determined from the 1998–99 IPEDS Institutional Characteristics file; some errors in this classification were uncovered when institutions were contacted.

<sup>2</sup>Based on the 1998–99 IPEDS Institutional Characteristics file.

<sup>3</sup>During 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 was increased by one.

NOTE: Institutions that offer baccalaureate degrees and higher were classified as either low or high education depending on the percentage of baccalaureate degrees awarded in education (based on the 1996-97 IPEDS completions file). Low education consisted of institutions in the bottom 80 percent of the distribution, while high education was the upper 20 percent. Details may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).



were sampled systematically on a flow basis according to fixed inclusion probabilities. These inclusion probabilities were designed to provide overall stratum sample sizes that would equal in expectation the previously determined target stratum sample sizes.

Summary statistics from the final individual level sample are given in table 2.2. From this table, it can be seen that of the final NPSAS:2000 sample size of about 70,230 individuals, about 16,620 of them had been identified as potential baccalaureates by the institutions. Marginal counts and percentages for the student sample by institutional control, institutional level, and institutional sector (the cross of institutional control by institutional level) are also provided.

#### **2.1.4 The B&B:2000/01 Sample**

The third stage of the B&B:2000/01 sample was a subsample of potential baccalaureate recipients who were CATI nonrespondents in NPSAS:2000. This subsample from NPSAS nonrespondents allows the B&B:2000/01 final sample to be more representative of the target population of students receiving baccalaureate degrees. To collect the subsample, a sampling frame was designed, subsampling strata defined, sample allocations within strata determined, and then the stratified sample was selected.

**The sampling frame and sample size.** The B&B:2000/01 sample restricted the NPSAS:2000 target population to individuals who received baccalaureate degrees. To find these individuals, institutions identified potential baccalaureate recipients prior to the start of the second stage of sampling. By the completion of the second NPSAS:2000 stage, two relevant types of students could be identified: (1) individuals confirmed as receiving a degree by the NPSAS:2000 CATI interview, and (2) individuals who remained potential baccalaureates due to CATI nonresponse.

More explicitly, the confirmed baccalaureate recipients were obtained as follows. At the completion of NPSAS:2000 sampling, before CATI there were about 16,620 undergraduate students sampled as potential baccalaureate recipients. About 11,340 completed the NPSAS:2000 CATI interview, of which about 9,830 were confirmed as receiving a baccalaureate degree during the 1999–2000 academic year.<sup>7</sup> Of the about 40,980 other undergraduates sampled, about 24,620 completed the CATI interview. Of these, about 490 received a baccalaureate degree in the time frame to become eligible for the B&B:2000/01 study. Of the graduate and first-professional students, about 80 reported in NPSAS:2000 receiving a baccalaureate degree during the 1999–2000 academic year. Therefore, there were about 10,400 confirmed baccalaureate recipients for the B&B:2000/01 subsample (9,830 + 490 + 80 = 10,400).

---

<sup>7</sup>Values presented in this paragraph can be found in table 4-21 of the NPSAS:2000 Methodology Report or the NPSAS:2000 data file.

**Table 2.2. Initial classification of NPSAS:2000 student sample, by type of institution and student stratum**

Institution type	Total sample <sup>1</sup>		Student sampling stratum <sup>2</sup>							
	Number	Percent	Baccalaureate sample <sup>3</sup>		Other under-graduate sample		Graduate sample <sup>3</sup>		First-professional sample	
			Number	Percent	Number	Percent	Number	Percent	Number	Percent
<b>All institutions</b>	70,230	100	16,620	100	40,980	100	11,280	100	1,350	100
<b>Institutional level</b>										
Less-than-2-year	6,670	10	†	†	6,670	16	†	†	†	†
2-year	13,240	19	†	†	13,240	32	†	†	†	†
4-year non-doctorate-granting	18,750	27	6,650	40	9,820	24	2,290	20	†	†
4-year doctorate-granting	31,570	45	9,980	60	11,250	28	8,990	80	1,350	100
<b>Institutional control</b>										
Public	43,750	62	10,750	65	25,970	63	6,540	58	490	36
Private not-for-profit	19,370	28	5,630	34	8,470	21	4,413	39	860	64
Private for-profit	7,110	10	250	2	6,540	16	330	30	†	†
<b>Institutional sector</b>										
Public less-than-2-year	1,530	2	†	†	1,530	4	†	†	†	†
Public 2-year	10,660	15	†	†	10,660	26	†	†	†	†
Public 4-year non-doctorate-granting	9,880	14	3,460	21	5,210	13	1,210	11	†	†
Public 4-year doctorate-granting	21,670	31	7,280	44	8,580	21	5,330	47	490	36
Private not-for-profit 2-year or less	1,840	3	†	†	1,840	5	†	†	†	†
Private not-for-profit 4-year non-doctorate-granting	8,010	11	3,030	18	4,040	10	930	8	†	†
Private not-for-profit 4-year doctorate-granting	9,530	14	2,600	16	2,590	6	3,480	31	860	64
Private for-profit less-than-2-year	4,520	6	†	†	4,520	11	†	†	†	†
Private for-profit 2-year or more	2,590	4	250	2	2,010	5	330	3	†	†

†Not applicable.

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

<sup>2</sup> As 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.

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

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

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

For the about 5,280 potential baccalaureate recipients who did not complete a CATI interview during NPSAS:2000, about 3,980 were still classified as potential baccalaureate recipients after NPSAS:2000 abstraction of student records via computer-assisted data entry (CADE). Of these, about 3,520 had complete CADE information. After consideration of various sampling options and budgetary constraints, it was decided to use the about 3,520 nonresponding potential baccalaureate CADE respondents and the about 10,400 confirmed baccalaureates as the sampling frame for the B&B:2000/01 study. Of these nearly 13,920 individuals, all confirmed baccalaureates were sampled, as well as about 1,300 of the approximately 3,520 nonresponding potential baccalaureate recipients. This sample size (10,400 + 1,300 = 11,700 initial sample size) was determined to provide adequate representation of baccalaureate recipients and to provide a final sample of approximately 10,000 baccalaureate respondents in the final B&B:2000/01 sample.

**Sample strata.** Strata for the sampling of potential baccalaureate recipients were formed by crossing the levels of institutional control, highest level of offering (doctorate-granting or not), and nonrespondent group, resulting in 25 non-empty strata. The first two of these variables were available from the first-stage NPSAS:2000 institution level sample, while the nonrespondent group was created specifically for this purpose. The stratification by institutional control and highest level of offering was done to preserve the sample distribution by type of institution found in the NPSAS sample.

The nonrespondent group indicator was a five-level variable created to categorize individual NPSAS:2000 CATI response status. Its levels (shown in table 2.3) were defined as those expected to be most predictive of our ability to locate and interview individuals for the B&B:2000/01 sample.

**Table 2.3. Expected response rates of potential baccalaureate recipients to the follow-up study, by NPSAS:2000 nonresponse status**

NPSAS:2000 nonrespondent status	Number of individuals	Percent of nonrespondents	Percent expected response rate
Total	3,980	100	40
Located, nonrefusal	200	5	75
Located, refusal	1,370	34	40
Not located, no tracing	330	8	60
Not located, post-CATI tracing	1,590	40	25
Not located after field tracing, or hostile refusal	500	12	15

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

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

**Sample stratum sample sizes.** Stratum sample sizes of potential baccalaureate recipients were obtained by first determining the sample allocation that maximized the expected response rate subject to the constraint that the unequal weighting effect not exceed 3.0 for any of

the strata defined by institutional control and highest level of offering. The expected response rates required for the optimization were determined through a combination of previous study experience and the results of the B&B:2000/01 field test. For the base-year nonrespondents, these expected response rates are provided in table 2.3.

Given the obtained optimal stratum allocations, final strata allocations were found by increasing the sampling rates for the strata with higher expected response rates. Simulation was then used to verify that the design did not lead to excessive unequal weighting effects. The simulation evaluated four sampling designs to assess their influence on unequal weighting effects. These four designs were the defined two sample allocations (optimal and implemented), and two sampling methods (stratified simple random sampling or stratified probability proportional to size (pps) sampling) with the measure of size equal to the NPSAS:2000 study weight. In each simulation, a sample of about 1,300 CATI nonrespondents was selected and unequal weighting effects were computed, both overall and for the five non-empty levels of the cross of institutional control and highest level of offering. The results of these simulations are given in table 2.4, and demonstrate reduced unequal weighting effects from sampling with probabilities proportional to NPSAS:2000 study weights, as well as similar unequal weighting effects for the optimal and modified sample allocations. For this reason, the modified allocation was used and stratified pps sampling was conducted.

**Table 2.4. Simulation of unequal weighting effects of four sampling designs**

Type of four-year institution	Optimum allocation		Implemented allocation	
	Stratified simple random sample	Stratified probability minimum replacement	Stratified simple random sample	Stratified probability minimum replacement
<b>All institutions</b>	2.1	1.9	2.1	1.9
<b>Institutional sector</b>				
Public non-doctorate-granting	2.3	2.0	2.1	2.0
Public doctorate-granting	2.0	1.8	2.1	1.9
Private not-for-profit non-doctorate-granting	2.0	1.9	2.0	1.9
Private not-for-profit doctorate-granting	2.0	1.9	2.0	1.9
Private for-profit	1.9	1.5	1.6	1.5

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

The final sample allocation and sampling rates are presented in table 2.5. Sampling rates were lower in the hard to interview strata (strata 16–25 in the table) as compared to the easier to interview strata (1–15), although a number of these strata still have relatively large sample sizes.

**Within stratum sampling design.** Potential baccalaureate recipients were sampled using a stratified pps design, in which each individual’s measure of size was determined to be the estimate of their NPSAS:2000 study weight at the time of sample selection. To avoid multiple selection of any individual, an iterative method to identify certainty selections was used to select the remainder of the sample using a probability minimum replacement (pmr) algorithm (Chromy, 1979).<sup>8</sup> This resulted in partitioning the individuals in each stratum into two groups: individuals sampled with certainty and individuals sampled with probabilities proportional to NPSAS:2000 study weights.

<sup>8</sup> Chromy 1979.

**Table 2.5. Sample strata allocation and sampling rates among potential baccalaureate recipients**

Stratum	NPSAS:2000 tracing and response status	Type of institution	Number eligible	Sample number	Sampling rate
1	Located, nonrefusal	Public non-doctorate-granting	30	20	0.71
2		Public doctorate-granting	80	60	0.74
3		Private not-for-profit non-doctorate-granting	30	20	0.74
4		Private not-for-profit doctorate-granting	30	20	0.74
5		Private for-profit <sup>1</sup>	10	10	0.67
6	Located, refusal	Public non-doctorate-granting	240	90	0.37
7		Public doctorate-granting	590	220	0.37
8		Private not-for-profit non-doctorate-granting	210	80	0.37
9		Private not-for-profit doctorate-granting	240	90	0.37
10		Private for-profit <sup>1</sup>	20	10	0.39
11	Not located, no tracing	Public non-doctorate-granting	40	30	0.71
12		Public doctorate-granting	100	80	0.74
13		Private not-for-profit non-doctorate-granting	30	20	0.74
14		Private not-for-profit doctorate-granting	40	30	0.74
15		Private for-profit <sup>1</sup>	10	10	0.78
16	Not located, post-CATI tracing	Public non-doctorate-granting	260	80	0.29
17		Public doctorate-granting	640	190	0.29
18		Private not-for-profit non-doctorate-granting	220	70	0.29
19		Private not-for-profit doctorate-granting	220	60	0.29
20		Private for-profit <sup>1</sup>	50	10	0.30
21	Not located after field tracing, or hostile refusal	Public non-doctorate-granting	90	30	0.29
22		Public doctorate-granting	190	60	0.30
23		Private not-for-profit non-doctorate-granting	80	20	0.30
24		Private not-for-profit doctorate-granting	90	30	0.29
25		Private for profit <sup>1</sup>	10	#	0.33

# Rounds to zero. All zeros provided in this table are actual values.

<sup>1</sup>These strata are collapsed over doctorate and nondoctorate granting institutions.

NOTE: Sampling rates were based on unrounded values but have been rounded for this table. Details may not sum to total due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01)

The iterative sampling method was implemented as follows. Label the  $n_h$  individuals in stratum  $h$  from 1 to  $n_h$ , let set  $A_h$  contain all individuals selected with certainty, and  $A'_h$  contain the rest. Let  $f_h(i)$  be the frequency of selection of individual  $i$  for a pmr design based on  $A_h$  and  $A'_h$ , let  $n(A_h)$  be the size of set  $A_h$ , and let  $S_i$  be the best available NPSAS:2000 study weight estimate for individual  $i$  at the time when the sample was selected. Begin by setting  $A_h = \{1, \dots, n_h\}$ ,  $A'_h$  as the empty set, and iterate as follows:

- 1) For each  $i$  in  $A_h$  find its expected frequency of selection:

$$E[f_h(i)] = \frac{n(A_h)S_i}{\sum_{j \in A_h} S_j}.$$

- 2) For any  $i$  in  $A_h$  with  $E[f_h(i)] \geq 1$ , transfer  $i$  from  $A_h$  to  $A'_h$ .

Iterations continued until  $E[f_h(i)] \leq 1$  for all  $i$  in  $A_h$ . A without-replacement pps design was implemented based on the final  $A_h$ , and  $A'_h$ .

## **2.2 Instrument Design**

Data for the B&B:2000/01 longitudinal cohort were collected by CATI. In preparation for the development of the CATI instrument, a comprehensive set of data elements was developed by reviewing the data elements used for the B&B:1993/94 cohort, the relationship of those elements to the NPSAS:2000 data elements, and relevance to current research and policy issues. A preliminary set of B&B:2000/01 data elements was refined with input from the study's Technical Review Panel (see appendix A for a list of members), as well as from NCES and other U.S. Department of Education staff. The final set of data elements, presented in appendix B, was then reviewed by the Office of Management and Budget (OMB) before the start of data collection.

Based on the set of data elements, the CATI instrument was structured by identifying section topics and determining the progression of items within sections. Individual items were designed from the data elements with several goals in mind:

- using NPSAS:2000 and B&B:1993/94 items when feasible,
- ensuring consistency with NPSAS:2000 and B&B:1993/94 items when items were not identical, and
- identifying and preparing wording for item verifications and probes as necessary.

These items were analyzed during the field test and modifications were made where necessary. Facsimile instruments may be found in appendix C.

Instrument sections were reviewed on a flow basis by NCES as well as by selected contractor and subcontractor staff. As depicted in figure 2.2, the first section determined eligibility for sample members who were nonrespondents in NPSAS:2000. The following sections collected information on postsecondary enrollment since high school completion, respondent demographics, post-baccalaureate education, employment, and teaching experiences.

To minimize the interview burden on respondents, the CATI instrument used existing data whenever feasible. Base year data from the NPSAS:2000 interview were preloaded into the CATI interview, dictating the flow of many portions of the interview. Certain questions were asked only if the data were missing from the prior interview.

The CATI interviews were programmed using Computer-Assisted Survey Execution System (CASES) 4.3 software.<sup>9</sup> In addition to a sample-member locating and contacting component, and informed consent screens employed before the start of the interview, the CATI system presented interviewers with screens of questions to be asked of the respondents. The software guided the interviewer and respondent through the interview, automatically skipping questions not applicable based on prior response patterns. Wording for probing and verification was suggested when a respondent provided a response that was out of range for a given item. As the CATI instrument was being designed and programmed, instrument documentation was

---

<sup>9</sup> Computer-Assisted Survey Execution System, Version 3.4. University of California at Berkeley, CA.

entered into an integrated data dictionary system (DDS), which subsequently enabled users to produce deliverable data files with CATI variable documentation.

To better represent respondent groups that are particularly difficult to interview, such as CATI nonrefusals and sample members with a limited proficiency in English severe enough to preclude completing the entire interview even with assistance from a bilingual interviewer, an abbreviated instrument was developed that focused on several key items about the respondent's post-baccalaureate enrollment and work experiences. A copy of this abbreviated instrument is presented in appendix C. Use of an abbreviated instrument reduces the potential for nonresponse bias for questions on the abbreviated instrument. There remains, however, a potential for nonresponse bias for questions on the full CATI instrument that were not on the abbreviated instrument. An analysis of this potential bias can be found in section 6.5.5.

## **2.3 Data Collection Design**

The B&B:2000/01 data collection design involved tracing sample members to their current location, then conducting a computer-assisted telephone interview with them about their experiences since the NPSAS:2000 interview approximately one year earlier. Data collection activities are shown in figure 2.3 and included locating sample members, notification letter mailings, intensive tracing procedures, and interviewing.

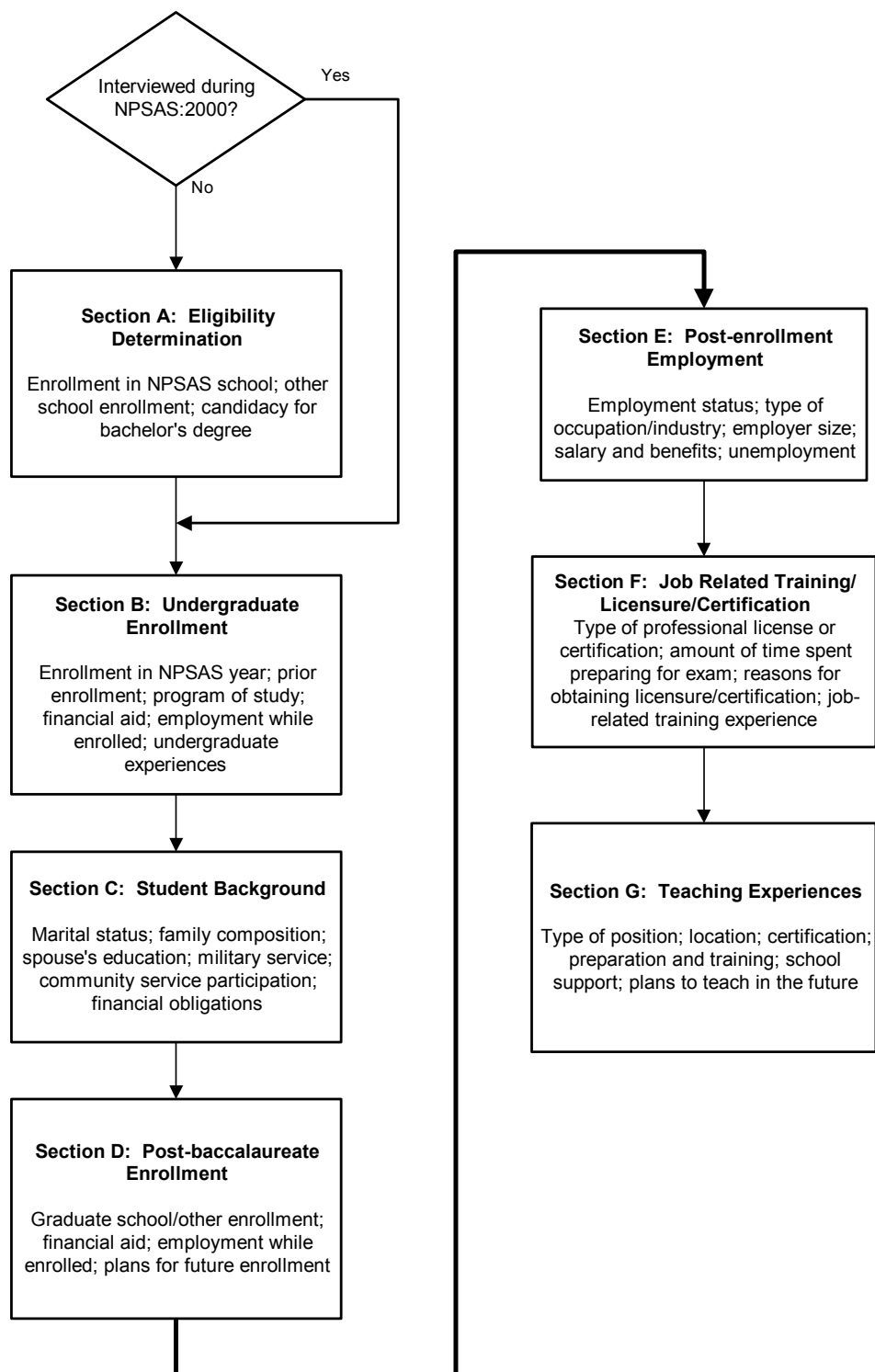
### **2.3.1 Pre-CATI Locating**

Before the start of interviewing, pre-CATI locating activities were employed to update address and telephone information for the sample members. This was a complex task. Initially, information received from NPSAS:2000 pertinent to this study was merged with the B&B:2000/01 locator database. This information was sent in batch mode first to the U.S. Postal Service National Change of Address (NCOA) system and then to Telematch in May 2001. These services provided updated address and telephone number information, respectively. Following the first round of NCOA and Telematch batch processes, the locator database was updated with the new information.

### **2.3.2 Student Notification Letter Mailing**

After addresses were updated, a notification mailing was sent to all sample members 1 week before data collection started to inform them of the study and their rights as participants. The mailing also gave sample members the opportunity to return an address update sheet. Samples of the lead letter, address update sheet, and information leaflet are included in appendix D. All locating information obtained from the student mailing was entered into the B&B:2000/01 locator database.

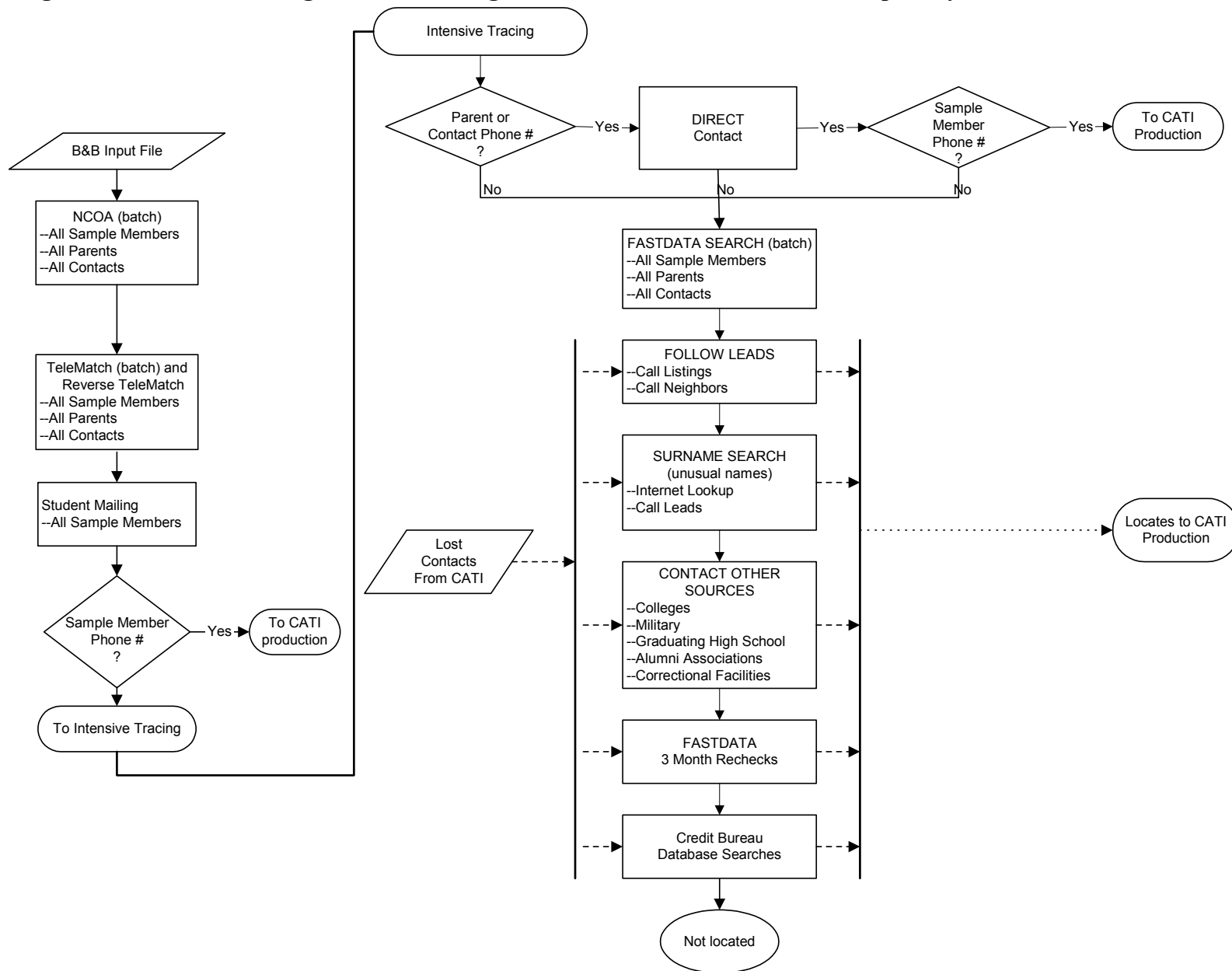
Figure 2.2. Structure and flow of B&B: 2000/01 follow-up student CATI



SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).



**Figure 2.3. Flow of locating and interviewing activities for B&B:2000/01 follow-up study**



SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000-2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

### **2.3.3 CATI Locating**

Locating and tracing activities took place concurrently with efforts to gain cooperation from sample members. When assigned a case, the telephone interviewer called the telephone number designated by CATI 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 did not provide the information needed, the interviewer initiated tracing procedures, using all other available information for other contact persons in an attempt to locate the student. When all tracing options available to the interviewer were exhausted, the case was assigned to RTI's Tracing Operations Unit (TOPS) for intensive tracing.

### **2.3.4 Intensive Tracing Efforts**

TOPS had access to both proprietary and public-domain locating databases. These provided real-time access to several consumer databases containing current address and phone listings for the majority of consumers with a credit history. In addition to the proprietary databases, TOPS had access to various other information sources, such as data miners, commercial list houses, and NCOA via a leased line. These sources provided the following searches: name, address, neighbor, business, phone matching searches, and status (decedent, incarcerated, incapacitated, or military personnel). TOPS employed these various information sources to locate respondents.

A two-tiered intensive tracing plan was used to locate B&B:2000/01 sample members. The first tier involved identifying sample members with Social Security numbers (SSNs) and processing that information through consumer database searches. If a search generated a new telephone number, that case was sent back to CATI for telephone interviewing. If a new address was generated but a new telephone number was not, tracers called directory assistance or accessed other databases to obtain telephone numbers for CATI. 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. The second tier of tracing activities included the following:

- checking directory assistance for telephone listings at various addresses;
- using electronic reverse-match databases to obtain the names and telephone numbers of neighbors, and then calling the neighbors;
- 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;

- contacting the current or last known residential sources such as the neighbors, landlords, current residents, tax assessors, realtors, and other business establishments related to previous addresses associated with the sample member;
- calling colleges, military establishments, and correctional facilities to follow up on leads generated from other sources; and
- the use of various tracing web sites.

Tracers checked new leads produced by these tracing steps to confirm the addresses and telephone numbers for the sample members. When the information was confirmed, that case was returned to CATI for telephone interviewing. Cases that could not be located (e.g., there were no working telephone numbers or numbers for relevant neighborhood sources were unpublished) were reviewed by supervisors, and (if necessary) were finalized as unlocatable.

### **2.3.5 Interviewing**

**Training of interviewers.** The interviewer training program was designed to maximize the trainees' active participation. Training manuals included a training guide, an interviewer's manual, and a question-by-question specification manual. Training for telephone interviewers and supervisors was conducted during the first week in July in several sessions, totaling 14 hours, and consisted of lectures, demonstrations, and hands-on practice exercises with the instrument and online coding modules. Trainees were introduced to the procedural aspects of data collection for B&B:2000/01 and were given a thorough review of the instrument. Interviewers were also trained in techniques for gaining cooperation with sample members, parents, and other contacts, as well as techniques for addressing the concerns of reluctant participants and avoiding refusals. A copy of the training agenda and the table of contents from the training manual are located in appendix E.

**Telephone interviewing.** CATI locating and interviewing were conducted from July 9, 2001 through November 20, 2001. CATI procedures included attempts to locate, gain cooperation from, and interview sample members. For NPSAS:2000 nonrespondents, NPSAS and B&B eligibility determination was also necessary. Locating information gleaned from the pre-CATI locating sources described above was preloaded for each case. Additionally, information previously collected through NPSAS:2000 was preloaded to personalize interviews and reduce respondent burden.

An automated call scheduler assigned cases to interviewers based on time of day, day of week, existence of precise appointments, and type of case. Case assignment was designed to maximize the likelihood of contacting and interviewing sample members, and cases were assigned to various queues for this purpose. Some of the queues included new cases, Spanish-language cases, initial refusals, and various appointment queues (appointments set by the sample member, appointments suggested by locator sources, and appointments for cases that were initial refusals).

For each case, a calling roster prioritized the names and telephone numbers for the interviewers to call. The roster included address information provided by schools and students. For example, addresses might have been permanent or local student addresses, parent addresses, and the addresses of other contacts from the NPSAS:2000 study. Up to six roster lines were preloaded with contact information. New roster lines were added as the result of CATI tracing and intensive tracing efforts.

Once located, some sample members required special treatment. To gain cooperation from those who initially refused to participate (including locator sources who acted as “gatekeepers” to prevent access to the sample member), certain interviewers were trained in refusal conversion techniques. Sample members and their locator sources who spoke only Spanish were assigned to bilingual interviewers.

## **2.4 Data Files**

The final B&B:2000/01 data files were prepared in accordance with NCES-specified ECB format guidelines. Activities important to ensuring quality across data file preparation methods are described below.

### **2.4.1 Data Availability Throughout Data Collection Period**

During data collection, RTI updated master data files containing completed case data. A master data file structure was retained across all CATI applications to concatenate data into a single data set containing all respondents. CATI data were extracted periodically to facilitate advance preparation of variable documentation. This extraction also enabled review of frequency distributions and data analysis while data were being collected.

### **2.4.2 Data Editing**

Data became available for editing during the data collection period, which allowed for feedback to the project staff on data quality as well as more accurate analysis of response data. As a quality check, the original skip logic was recreated to ensure that respondents followed the appropriate path within the CATI instrument. These edit checks were important for correcting errant paths that the interviewer may have followed but later corrected. Edit checks ensured that, for particular data elements, responses occurred logically. This process also allowed evaluation of the other case: when questions that the respondent should have been asked were missed. Reserve codes indicated instances in which raw data were updated to reflect the proper logical path. The data editing process is discussed in detail in chapter 5. During data collection, interviewing staff were also able to notify project staff of CATI irregularities via “problem sheets,” so problems in instrument logic could be corrected.

## 2.5 Integrated Management System

All aspects of the study were under the control of an integrated management system (IMS). The IMS consisted of several components or modules:

- Management
- Receipt Control System (RCS)
- CATI

This modular structure allowed for the streamlining of related tasks and resulted in a centralized, easily accessible repository for project data and documents. These modules are described in more detail below.

The Management module of the IMS contained tools and strategies to assist project staff and the NCES project officer in managing the study. All information pertinent to the study could be found via the World Wide Web in a secure desktop environment, including schedules, monthly progress reports, project plans and specifications, information related to the Technical Review Panel (TRP), and project deliverables. Also available in the management module were the latest version of the CATI instrument, daily RCS module status reports, and daily data collection reports.

The RCS module monitored activities related to data collection, including tracing and locating, thereby enabling project staff to perform stage-specific activities, track case status, identify problems early, and implement solutions effectively. Several applications used the RCS's locator data for daily tasks: (1) the mailout program produced mailings to parent/contacts and sample members, (2) the query system enabled administrators to review the locator information and status for a particular case, and (3) the mail return system enabled project staff to update the locator database as lead mailings or reply sheets were returned. The RCS also interacted with the tracing operation system database, sending locator data between the two systems as necessary.

The CATI module managed development of the CATI instrument within the data dictionary system (DDS). The DDS consisted of a set of linked relational files and associated utilities for developing and documenting the instrument. Developing the CATI instrument with the DDS ensured that all variables were linked to their item/screen wording and were thoroughly documented. Also included within the CATI module was online coding software ("user exits") that collected detail on schools attended, enrollment, industry, occupation, and major field of study.



# Chapter 3

## Data Collection Procedures

---

This chapter evaluates the procedures used for the B&B:2000/01 study. Three general areas were evaluated: respondent locating, obtaining respondent interviews, and the efforts expended in obtaining the interview.

### 3.1 Locating Outcomes

One of the biggest challenges facing data collection was the aggressive 5-month schedule for completing the targeted 10,000 interviews. The effort required the use of a number of varied tracing approaches and continual refinement of these techniques to locate sample members in sufficient time to interview them. The level of time and effort required to complete the locating and interviewing steps with sample members varied considerably. Some sample members were reached and interviewed on the first contact attempt. Others required considerable tracing (contacting parents, former roommates, etc.) before they were successfully located and interviewed.

A high percentage (86 percent) of sample members were located and interviewed as part of B&B:2000/01. This is at least partially because the B&B:2000/01 study is a relatively quick follow-up. Base-year data were collected for respondents only 1 year earlier during NPSAS:2000, which aids the success of locating efforts. Even for this highly mobile population of recent college graduates, locating is much easier 1 year after initial contact than in a later follow-up.

#### 3.1.1 Student Locating and Interviewing

Figure 3.1 presents a schematic of the outcomes of student locating, interviewing, and related case-resolution activities. Data were collected almost exclusively by CATI, with about 200 cases completed using hard-copy questionnaires administered by field locating staff. The small set of approximately 30 field locators was used near the end of data collection to locate cases in which centralized tracing efforts failed to turn up a valid address or telephone number for the sample member. Field locators were persons in close regional proximity to the last known address of the sample member, and who had access to additional sources of information to assist in locating hard-to-find cases.

Figure 3.1 summarizes the attempts that were made to locate the approximately 11,700 sample members. Overall, about 10,690 of these cases (91 percent) were located, about 760 (7 percent) were *not* located, about 190 (2 percent) were considered “exclusions,”<sup>10</sup> and

---

<sup>10</sup>B&B exclusion cases were considered “out-of-scope” for locating and interviewing operations. Among the 190 B&B:2000/01 sample members classified as exclusions, 170 were out of the country for the duration of the data collection period, 10 sample members were deceased, and 10 were incapacitated and unable to complete the survey.

approximately 70 (<1 percent) were ineligible<sup>11</sup> for the study based on their responses to the eligibility questions in the instrument.

Student interviewing results are also shown in figure 3.1. A total of about 10,030 (of the approximately 11,520 remaining cases after removing the exclusions) were interviewed. Many completed the entire interview (about 9,650), while approximately 380 completed a partial or abbreviated interview. A small number of the interviews (about 10) were classified as partial because the respondent broke off after completing part of the interview. A break-off represented an explicit or implicit refusal, but such cases could not be converted or recontacted to complete the interview by the end of the data collection period. The remaining number of these interviews (about 370), however, resulted from the administration of an abbreviated interview that consisted of a minimal set of questions from the full interview. Over half of this group represented interviews with Spanish-speaking respondents.

A total of about 1,420 cases were not interviewed. These sample members included approximately 760 who were not located. In addition, nearly 660 potentially eligible students were located but were not interviewed. Of these, about 530 were explicit final refusals for which subsequent attempts at interviewing were determined to be infeasible or unwise. Cases not interviewed also included about 130 sample members for whom the data collection period expired before they could complete the interview; such cases clearly reflect, at least in part, the constricted data collection period.<sup>12</sup>

An unweighted CATI response rate for B&B:2000/01 can be calculated as the number of respondents interviewed divided by the initial sample size minus the ineligibles: student CATI response rate =  $10,030 / (11,700 - 70) = 86$  percent. The weighted overall CATI response rate was 75 percent.

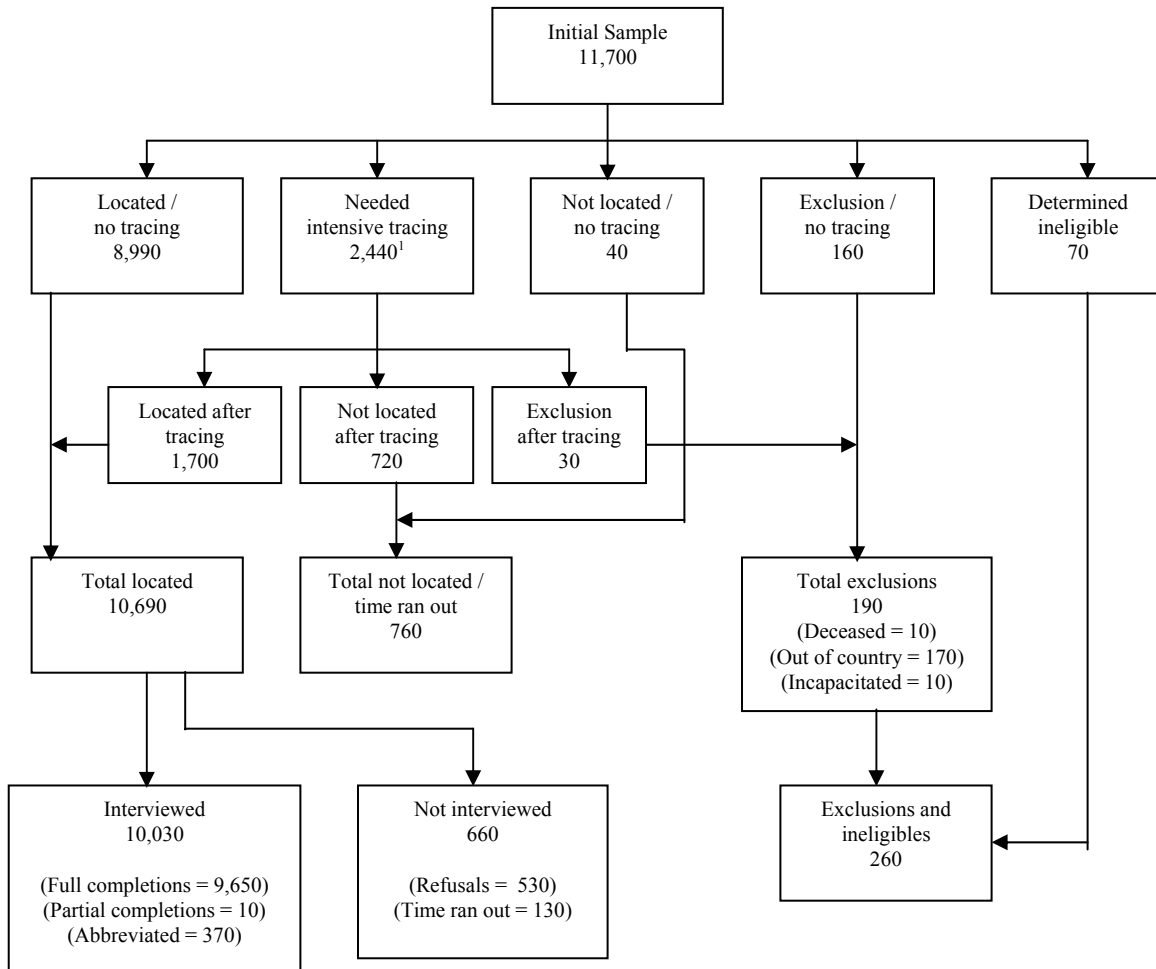
---

<sup>11</sup>In NPSAS, persons could have been CADE eligible but CATI ineligible if they told us in the interview that they did not attend the NPSAS school at any time during the NPSAS year (7/99–6/00) or had not received their baccalaureate degree from the NPSAS institution. CADE eligibility was determined based on the records from the institution—CATI eligibility was determined based on responses in CATI.

<sup>12</sup>This group likely contained an unknown number of implicit refusal cases—i.e., individuals who after first contact used answering machines or friends/relatives as gatekeepers, as well as those who continued to make (and then break) appointments for an interview in the future.



Figure 3.1. B&B:2000/01 results of locating and interviewing activities



<sup>1</sup>Includes all cases sent to RTI's Tracing Operations (TOPS) for intensive locating efforts. Overall, 130 cases required pre-CATI intensive tracing only, 1,200 required post-CATI intensive tracing only, and 1,100 required both pre- and post-CATI intensive tracing.

NOTE: Final sample including interviewed and exclusions = 11,630. Details may not sum to total due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

Locating and interviewing rates varied by base year respondent status. Table 3.1 provides results for B&B:2000/01 student locating and interviewing (for those located) by respondent status in the base year study (NPSAS:2000).<sup>13</sup> In terms of locating, nearly 96 percent of the NPSAS:2000 respondents were located, compared to 73 percent of NPSAS:2000 nonrespondents. There were also differences in interview rates among the two groups. Among base year respondents, 94 percent of the located sample members successfully completed interviews during the B&B:2000/01 follow-up study, compared to 69 percent of located NPSAS:2000 nonrespondents.

**Table 3.1. B&B:2000/01 student locating and interview results, by respondent status in NPSAS:2000**

NPSAS:2000 respondent status	Total	Located		Interviewed, given locate	
		Number	Percent	Number	Percent
Total	11,630	10,870	94	10,030	92
NPSAS:2000 respondent	10,400	9,970	96	9,400	94
NPSAS:2000 nonrespondent	1,240	910	73	630	69

NOTE: Statistics exclude 70 B&B-ineligible sample members (as determined in CATI). Details may not sum to total due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

### 3.1.2 Pre-data Collection Tracing and Locating Efforts

Prior to the start of data collection, RTI took a number of steps to ensure that the locating information being fed into CATI was as accurate as possible. The pre-CATI tracing and locating efforts involved batch tracing for all cases, using database matching with National Change of Address (NCOA) and Telematch databases, pre-data collection intensive locating by the tracing operations unit, and mailing of advance letters with requests for sample members to return locating update sheets with new address and telephone information. These steps helped get data collection off to a fast start.

**Batch matching to NCOA database.** One month prior to the start of data collection, all of the B&B:2000/01 cases, including subject and contact information, were sent to the NCOA for address updating. The NCOA file is a consolidated file that contains approximately 108 million permanent change-of-address records filed with the U.S. Postal Service, retained on the file for a 3-year period from the move-effective date.

Table 3.2 provides the locate and interview rates for cases sent to NCOA for address updating. Of the nearly 11,490 eligible cases sent to NCOA, about 1,780 (16 percent) were returned with new or confirmed address information. For cases in which locating information was confirmed, 92 percent resulted in contact with the sample member. Of these, 93 percent resulted in a completed interview. For those cases where NCOA did not return a match, 94 percent of the cases were located, of which 92 percent completed the interview.

<sup>13</sup>The statistics in table 3.1 exclude 70 B&B:2000/01 ineligible sample members determined during CATI interviewing and 190 exclusions; they do not exclude any potential ineligibles likely to be part of the unlocatables.

**Table 3.2. B&B:2000/01 locate and interview rates, by pre-CATI NCOA processing**

Student mailing status	Total	Contacted		Interviewed, given contact	
		Number	Percent	Number	Percent
Total	11,490	10,750	94	9,940	93
Confirmed/new information from NCOA	1,780	1,650	92	1,530	93
No new information from NCOA	9,710	9,110	94	8,410	92

NOTE: Statistics exclude 70 B&B-ineligible sample members (as determined in CATI) and 140 cases with insufficient address information to send to NCOA. Details may not sum to total due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

**Batch matching to Telematch database.** The second step in the pre-data collection batch processing involves sending all address information, including updates from NCOA, to Telematch for reverse telephone look-up services. Telematch provides a computerized residential telephone number look-up service using a database of more than 130 million U.S. telephone numbers. Table 3.3 provides the results of the batch matching to Telematch. Of the 11,490 eligible B&B:2000/01 records submitted, two-thirds (67 percent) were returned with new or confirmed telephone information.

**Table 3.3. B&B:2000/01 locate and interview rates, by pre-CATI Telematch processing**

Student mailing status	Total	Contacted		Interviewed, given contact	
		Number	Percent	Number	Percent
Total	11,490	10,750	94	9,940	93
Confirmed/new information from Telematch	7,670	7,340	96	6,850	93
No new information from Telematch	3,820	3,420	90	3,090	90

NOTE: Statistics exclude 70 B&B-ineligible sample members (as determined in CATI) and 140 cases with insufficient address information to send to Telematch. Details may not sum to total due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

**Intensive pre-data collection centralized tracing.** All NPSAS:2000 nonrespondent cases, as well as those cases without a valid address or telephone number, received intensive tracing. These are the cases believed to be the most difficult to locate at the outset of the study. As shown in table 3.4, about 1,240 cases were sent TOPS for advance tracing. Of those, 70 percent (about 910 cases) were ultimately located, and just over two-thirds of those (69 percent) completed the full, partial, or abbreviated hard-copy interview.

**Table 3.4. B&B:2000/01 contact and interview rates, by pre-CATI intensive tracing efforts**

Tracing status	Total	Contacted		Interviewed, given contact	
		Number	Percent	Number	Percent
Total	11,630	10,870	94	10,030	92
No intensive tracing required	10,400	9,970	96	9,400	94
Intensive tracing required	1,240	910	73	630	69

NOTE: Statistics exclude 70 B&B-ineligible sample members. Details may not sum to total due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

**Lead letter and locator mailing to students.** One week before the start of data collection, sample members were sent a letter and address update sheet. Each sample member was asked to review, correct, and return the sheet. Letters were mailed to about 11,630 sample members, and approximately 1,090 address update sheets with new or confirmed information were received (9 percent of the total sample).

Not surprisingly, sample member contact and interview rates varied considerably by whether a sample member returned the address update sheet (see table 3.5). While the return rate for these sheets was just over 9 percent, the contact and interview rates for those who did return the forms was nearly universal. More than 99 percent of those who returned the form completed the interview.

**Table 3.5. B&B:2000/01 locate and interview rates, by student return of address update form**

Student mailing status	Total	Contacted		Interviewed, given contact	
		Number	Percent	Number	Percent
Total	11,630	10,870	94	10,030	92
Returned update form	1,090	1,080	99	1,070	99
No update form returned	10,540	9,790	93	8,960	92

NOTE: Statistics exclude 70 B&B-ineligible sample members (as determined in CATI). Details may not sum to total due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

### 3.1.3 Tracing and Locating Efforts During Data Collection

Tracing during data collection took many forms. In addition to the contacting and locating done as part of the CATI contacting and interviewing processes, an array of other approaches was used to locate sample members and encourage them to participate in the survey. These efforts included centralized tracing by TOPS, in-person tracing by field locators, batch tracing (for a second time) to NCOA and Telematch, batch tracing to an additional vendor (TransUnion), mailout of a hard-copy questionnaire, and a postcard follow-up.

**Intensive locating during data collection.** Intensive tracing efforts were required for cases in which preloaded CATI locating information did not result in contact with the sample member. These cases were assigned to TOPS for intensive centralized tracing. The final locate and interview rates for cases requiring centralized tracing are provided in table 3.6. Of the approximately 11,630 eligible cases, nearly one in five (20 percent) required TOPS intensive tracing during data collection. Of these, 69 percent were located and, of those located, just under 92 percent were interviewed.

**Table 3.6. B&B:2000/01 contact and interview rates, by post-CATI intensive tracing efforts**

Tracing status	Total	Contacted		Interviewed, given contact	
		Number	Percent	Number	Percent
Total	11,630	10,870	94	10,030	92
No intensive tracing required	9,330	9,280	100	8,560	92
Intensive tracing required	2,310	1,600	69	1,470	92

NOTE: Statistics exclude 70 B&B-ineligible sample members (as determined in CATI). Details may not sum to total due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

**Source of locating information for completed interviews.** Locating students in a longitudinal study is a complex task, requiring multiple sources of information. Leads developed through one source may need to be verified using another data source or locating technique. For hard-to-locate sample members, generally no single source of information is adequate to achieve the level of locating required. Rather, a successful locating effort requires blending multiple sources of information. Table 3.7 provides an overview of the sources used during intensive tracing of the hard-to-reach B&B:2000/01 sample members. Note that although the table provides information on the number and percentage of sample members who were ultimately located when a particular source was used, most of the cases were traced using multiple sources.

**Table 3.7. B&B:2000/01 locate rates, by tracing source used during intensive tracing efforts**

Tracing source	Cases involving intensive tracing		
	Total	Individuals located	
		Number	Percent
Consumer database search	2,026	1,362	67
Directory assistance	1,995	1,330	67
Database–address search	1,756	1,145	65
Internet search	1,925	1,263	66
Database–reverse phone lookup	1,272	819	64
Database–name search	1,065	636	60
Database–neighbor search	406	234	56

NOTE: Most cases were traced using multiple sources, and therefore, the numbers and percentages are not mutually exclusive.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

A number of techniques proved effective in the locating effort. Contact was made with approximately two-thirds of the cases in which the following locating sources were used: consumer databases (67 percent), directory assistance (67 percent), searches of proprietary databases (65 percent), Internet searches (66 percent), and database searches using reverse telephone look-ups—that is, matching a telephone number to a known address (64 percent). Look-ups using just names or neighbor searches located over half of these sample members.<sup>14</sup>

**Database matching of nonrespondent cases.** To secure the highest possible response rate for B&B:2000/01, the address and telephone number information for cases that had not responded by October 1, 2001, were resubmitted to NCOA and Telematch to see if further updates could be obtained. Additionally, in the last month of data collection for B&B:2000/01, all nonrespondent cases were also processed against the TransUnion credit service database.

The results of the second batch processing through NCOA are shown in table 3.8. Of the approximately 3,400 nonrespondent cases sent to NCOA, only about 320 (about 10 percent) turned up with new or confirmed information. Of these, 68 percent were ultimately located, and 77 percent of located sample members completed the interview.

**Table 3.8. B&B:2000/01 locate and interview rates, by nonrespondent NCOA processing**

Student mailing status	Total	Contacted		Interviewed, given contact	
		Number	Percent	Number	Percent
Total	3,400	2,650	78	2,040	77
Confirmed/new information from NCOA	320	220	69	170	78
No new information from NCOA	3,080	2,430	79	1,860	77

NOTE: Statistics exclude 50 cases sent to NCOA but later determined to be ineligible. Details may not sum to total due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

The results of the nonrespondent processing through Telematch are shown on table 3.9. Over half (57 percent) of the nearly 3,190 cases submitted to Telematch produced either new or confirmed telephone numbers. Among the cases in which the locating information was updated, 83 percent were contacted, and of those contacted, 77 percent completed the interview.

<sup>14</sup> Neighbor searches involve the use of databases to identify the addresses and telephone numbers of properties or apartments located in close proximity to the sample member's last known address. The assumption is that current or former neighbors may be able to provide current locating information for the sample member.

**Table 3.9. B&B:2000/01 locate and interview rates, by nonrespondent Telematch processing**

Student mailing status	Total	Contacted		Interviewed, given contact	
		Number	Percent	Number	Percent
Total	3,190	2,440	77	1,860	76
Confirmed/new information from Telematch	1,830	1,510	83	1,160	77
No new information from Telematch	1,360	940	69	700	75

NOTE: Statistics exclude 50 cases sent to Telematch but later determined to be ineligible. Details may not sum to total due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

In the last month of data collection, RTI secured access to a new consumer database batch processing service offered by TranUnion. All nonrespondent cases as of October 15, 2001, were submitted for matching against this service to identify new leads or confirm information already obtained. A total of about 2,650 cases were submitted for processing, of which 98 percent were returned with either new information or a confirmation of an existing telephone number or address (see table 3.10). Among the cases returned with a match, 73 percent were located by study’s end. Seven of every 10 previously nonresponding sample members who could be contacted through these efforts completed the survey.

**Table 3.10. B&B:2000/01 locate and interview rates, by results of nonrespondent TransUnion processing**

Student mailing status	Total	Contacted		Interviewed, given contact	
		Number	Percent	Number	Percent
Total	2,650	1,910	72	1,326	70
Confirmed/new information from TransUnion	2,600	1,900	73	1,320	70
No new information from TransUnion	50	10	20	10	100

NOTE: Statistics exclude 50 cases sent to TransUnion but later determined to be ineligible. Details may not sum to total due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

**Field locating.** In an effort to boost response rates in the final month, 630 cases were sent to field locators, who encouraged located sample members either to complete the full interview by calling RTI’s CATI facilities or to complete a hard-copy<sup>15</sup> interview with the field locator. Of the nearly 630 cases sent to the field, about 400 sample members completed interviews, for an unweighted response rate of 64 percent (see table 3.11). Half of the completes (50 percent) were hard-copy completions obtained by the field staff. The other 50 percent were

<sup>15</sup> A hard-copy interview is one completed using a paper version of the abbreviated instrument.

completed by telephone after the field locator either provided the CATI facilities with new contact information or prompted the sample member to call in to complete the survey.

**Table 3.11. B&B:2000/01 locate and interview rates for field locator cases**

Case status	Number	Percent
Total cases	630	100
Complete-total	400	64
(Complete-by telephone)	200	32
(Complete-by abbreviated hard-copy questionnaire)	200	32
Located-time ran out	20	3
Located-refused	10	2
Located-confirmed exclusion	10	2
Located-confirmed ineligible	10	2
Not located	180	29

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

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

## 3.2 Interviewing Outcomes

### 3.2.1 Hard-copy Questionnaire Mailout and Postcard Follow-up

During the final month of data collection, hard-copy abbreviated instruments were mailed to all nonresponding sample members for whom a valid mailing address was available. Two weeks later, persons who had not yet responded were sent a postcard reminder. Both the mailing and the postcard informed the sample members that they could complete the hard-copy and return it in a business reply envelope (provided in each mailout packet) or call into a toll-free number at our CATI facilities to complete the interview.

Questionnaires were mailed to about 2,630 nonresponding sample members (see table 3.12). By the end of the study, completions were obtained from approximately 1,630 of these individuals (62 percent). Among these responding sample members, about 1,390 sample members (85 percent) completed the survey by telephone, while about 250 sample members (15 percent) completed and returned the hard-copy abbreviated questionnaires.

**Table 3.12. B&B:2000/01 locate and interview rates, by end of study hard-copy questionnaire mailout and postcard follow-up**

Student mailing status	Total	Contacted		Interviewed, given contact	
		Number	Percent	Number	Percent
Sent abbreviated hard-copy questionnaire	2,630	2,120	81	1,630	77
Sent postcard reminder	1,250	570	46	340	60

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

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).



### 3.2.2 Refusal Conversion Efforts

Refusal conversion procedures were used to gain cooperation from individuals who refused to participate when contacted by telephone interviewers. Refusals came not only from sample members, but also from spouses, housemates, parents, and other gatekeepers. When either a sample member or a gatekeeper refused to participate in the locating or interviewing effort, the case was referred to a specially trained refusal-conversion specialist in RTI's Telephone Survey Department. There were about 1,520 initial refusals among the B&B:2000/01 sample (13 percent of the eligible sample of approximately 11,630). Of the potential pool of eligible initial refusal cases, 64 percent were successfully converted and completed interviews.

**Table 3.13. B&B:2000/01 conversion of initial refusals, by respondent status in NPSAS:2000**

NPSAS:2000 field test respondent status	Number of initial refusals	Interviewed, given initial refusal	
		Number	Percent
Total	1,520	970	64
Respondent	1,270	890	70
Nonrespondent	260	80	31

NOTE: Statistics exclude 10 refusal cases found to be B&B ineligible (as determined in CATI). Details may not sum to total due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

The success of refusal conversion varied considerably according to the sample member's response status in the base year study (see table 3.13). Among respondents to the NPSAS:2000 interview, 70 percent of those who initially refused to be interviewed (or whose gatekeeper refused) ultimately completed the B&B:2000/01 interview. For base year nonrespondents who initially refused in B&B:2000/01, 31 percent were successfully converted.

### 3.3 Interview Burden, Time, and Effort

#### 3.3.1 Timing

This section reviews the effort and burden associated with the B&B:2000/01 student interview. Interview length was examined by considering the timing analysis statistics. This information is useful because it provides evidence that can reduce respondent burden, reduce data collection effort and cost, and improve data quality for future studies. The effort required to locate and interview sample members for the study was evaluated using the average interview time.

The CATI instrument development was embedded with time stamps at the start and end of the interview, as well as at the beginning and end of each interview screen, which could include up to eight related items. The time stamps measured the elapsed time to complete each segment of the interview. This enabled project staff to monitor the time required to complete specific interview items, the online coding programs, sections of the interview, and the interview as a whole.

Table 3.14 presents timing results for the B&B:2000/01 study cohort. The time (in minutes) needed to conduct a student interview is shown to respondents by interview section. Sections are listed in the table in the order in which they were presented to respondents. There were fluctuations in the number of cases completing each section for two reasons: (1) some respondents broke off the interview before completing all of the sections (partial interview), and (2) certain sections of the interview applied to selected groups of respondents (see figure 2.2), so timing results are presented for the overall cohort and by subgroup. For example, Section A was designed for base year nonrespondents, so the number of cases in that group was less than that for the rest of the instrument. Respondents who were currently teaching skipped Section E (on post-baccalaureate employment) and proceeded directly to the teaching section.

Overall average administration time to complete the interview was 19 minutes. Both respondents and nonrespondents to the NPSAS:2000 base year study took an average of approximately 19 minutes to complete the interview. As indicated in Table 3.14, base year respondents took longer to complete the sections on student background, post-baccalaureate enrollment, post-baccalaureate employment, training and certification, and teaching experience when compared to base year nonrespondents.

For respondents who had taught since graduating, the average interview time was 24 minutes, compared to 18 minutes for those who had not taught. Teachers took longer to complete many of the interview sections. These included sections on student background, post-baccalaureate enrollment, training and certification, and, as would be expected, teaching experience. On one section, post-baccalaureate employment, teachers took less time because many employment questions were skipped once teaching status had been confirmed.

### **3.3.2 Number of Calls and Response Rates**

Telephone interviewers made over 214,000 telephone calls during the data collection period, with an average of 18 calls per sample member.<sup>16</sup> An average of 17 calls per case was required to obtain the completed cases. Among the completed cases, nearly 44 percent were completed with fewer than 10 calls, an additional 41 percent required 10 to 29 calls, and just over 15 percent of the completed cases required 30 or more call attempts.

Of the total calls made, one in four resulted in contact with an individual. An answering machine was reached on 44 percent of the attempts, reflecting the heavy usage of such devices among this population. An answering machine was reached at least once (and often multiple times) for three of every four B&B sample members. The remaining 30 percent of calls made were other noncontacts (busy, ring/no answer, fax line, pager, etc.).

As might be expected, the number of completed interviews remained steady during data collection, and then completions tapered off towards the end as cases were harder to reach. Table 3.15 provides a summary of the proportion of total responses obtained for each week of data collection.

---

<sup>16</sup>These figures are based on calls made by telephone interviewers and exclude calls made by TOPS in the course of attempting to locate sample members.

**Table 3.14. Average minutes to complete B&B:2000/01 student interview, by interview section, NPSAS:2000 response status, and teaching status**

CATI section	Total		NPSAS:2000 nonrespondent		NPSAS:2000 respondent		B&B:2000/01 nonteachers		B&B:2000/01 teachers	
	Average time	Number of cases	Average time	Number of cases	Average time	Number of cases	Average time	Number of cases	Average time	Number of cases
Section A – eligibility (NPSAS nonrespondents)	0.42	540	0.42	540	†	9,120	0.41	450	0.43	90
Section B – enrollment history	6.16	9,650	6.09	540	6.16	9,110	6.19	7,640	6.05	2,010
Section C – student background	2.74	9,630	2.64	540	2.74	9,090	2.72	7,630	2.80	2,000
Section D – post-baccalaureate education	3.00	9,630	2.60	540	3.02	9,090	2.91	7,630	3.33	2,000
Section E – post-baccalaureate employment	4.20	9,620	4.02	540	4.21	9,090	4.46	7,620	3.20	2,000
Section F – training and certification	1.47	9,270	1.32	540	1.47	9,090	1.29	7,620	2.14	2,000
Section G – teaching experience	1.74	9,610	1.52	530	1.75	9,080	0.45	7,610	6.62	2,000
Total interview	19.29	9,620	18.51	540	19.33	9,080	18.05	7,620	24.01	2,000

† Not applicable.

NOTE: A section was considered complete if the amount of time to complete the section was greater than zero and the section completion flag was set. Section outliers were removed from the timing calculations (20 in section A, 10 in section B, 20 in section C, 10 in section D, 10 in section E, 0 in section F, and 20 in section G). Respondents may have completed one or more sections but not completed the entire interview. Only completed cases were included in the total interview timing results. Details may not sum to total due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

**Table 3.15. Proportion of total responses by time period (in weeks)**

Week	Weekly completes	Cumulative completes	Cumulative percent of total respondents
1	730	730	7
2	890	1,620	16
3	790	2,410	24
4	640	3,050	30
5	600	3,650	36
6	590	4,240	42
7	650	4,890	49
8	620	5,510	55
9	590	6,100	61
10	680	6,780	68
11	550	7,330	73
12	430	7,760	77
13	370	8,130	81
14	380	8,510	85
15	270	8,780	88
16	380	9,160	91
17	350	9,510	95
18	250	9,760	97
19	230	9,990	100
20	40	10,030	100

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

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

# Chapter 4

## Evaluation of Data Quality

---

A variety of methods were designed to assess the data quality of the B&B:2000/01 follow-up study. This chapter describes these methods and provides the results of them. Table 4.1 provides a summary of the types of data quality evaluation used in the study. Each will be described in the sections below.

**Table 4.1. Summary of B&B:2000/01 data quality evaluations**

Major area of evaluation	Evaluation approaches
<b>Instrument quality</b>	<ul style="list-style-type: none"> <li>• Analyze distribution of indeterminate responses (“refused” and “don’t know”) for interview sections and individual items.</li> <li>• Analyze the frequency of help text usage for each item.</li> </ul>
<b>CATI administration</b>	<ul style="list-style-type: none"> <li>• Analyze success and accuracy of online coding of industry/occupation and major course of study.</li> <li>• Debrief interviewers, refusal converters, bilingual interviewers, monitors, and supervisors in quality circle meetings.</li> <li>• Analyze silent monitoring quality control data, including accuracy of question delivery and data entry.</li> </ul>

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

### 4.1 Indeterminate Responses

Special keyed entry by interviewers (F3=don’t know or F4=refused) allowed the CATI interview to accommodate responses of “don’t know” and “refused” to every interview item. Refusal responses to interview questions were most common for items considered sensitive by respondents, while “don’t know” responses may have resulted from a number of potential circumstances. The most obvious reason a respondent will offer a “don’t know” response is that the answer is truly unknown or in some way inappropriate for the respondent. “Don’t know” responses may also be evoked when (1) question wording is not understood by the respondent (with no explanation by the interviewer), (2) the respondent hesitates to provide a “best guess” response (with insufficient prompting from the interviewer), or (3) a respondent implicitly refuses to answer a question. “Refused” and “don’t know” responses need to be reduced to the

greatest extent possible. These responses introduce indeterminacies in the data, and must be resolved by imputation or subsequently dealt with during analysis.

Overall item nonresponse rates were low, with only 6 items of 556 containing over 10 percent missing data. These items are shown in table 4.2 and are grouped by interview section. Item nonresponse rates were calculated based on the number of sample members for whom the item was applicable and asked. One item with a high rate of nonresponse pertained to income. Many respondents were reluctant to provide information about personal and family finances. Among those who were not reluctant, many simply did not know, perhaps because they were in new careers, recently changed jobs, or were in job transition. In addition, the items pertaining to the lifetime learning tax credit also garnered a high number of “don’t know” responses. These “don’t know” responses are most likely attributable to respondents’ unfamiliarity with the tax credit because of its relatively recent implementation.

**Table 4.2. B&B:2000/01 interview item nonresponse for items with more than 10 percent “don’t know” or “refused”**

CATI section and variable name	Label	Number asked	Percent don’t know	Percent refused	Combined percent
Student background					
C_INCS99	Spouse’s income in 2000	3,060	11	11	21
C_SPLON	Money borrowed in loans—not from family	2,340	11	1	12
Post-baccalaureate enrollment					
D_CREDIT	Claim lifetime learning tax next year	4,140	16	1	16
Post-baccalaureate employment					
E_SIZE	Number of employees at company	5,250	11	#	12
E_CURINC	Current annual income	9,990	7	6	13
E_TIME	Time scale of pay <sup>1</sup>	680	15	3	18

# Rounds to zero.

<sup>1</sup>Time scale of pay refers to per hour, week or month.

NOTE: Statistics are based on student sample members for whom specific items were applicable and asked. Items applicable to fewer than 100 sample members were excluded from consideration. Details may not sum to zero due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

## 4.2 Help Text

Online help text was available for every screen in the CATI instrument. Having additional information available at the touch of a key was beneficial to interviewers, particularly at the beginning of data collection, when the flow and content of the instrument were new to the data collection personnel. Online help text access served to immediately clarify response categories on item wording while interviewers were still on the telephone with the respondent. Help text screens displayed information designating to whom the item applied, the type of information requested in the item, and definitions of words or phrases in the item.

Counters were used to determine the number of times each help screen was accessed, making it possible to identify items that were confusing to interviewers or respondents.

Table 4.3 presents CATI items having the highest rates of help text usage,<sup>17</sup> along with their rates of indeterminacy. Overall, the rates of help text usage were very low. Two questions concerning the lifetime learning tax credit had a large number of accesses to help text, almost certainly because of student unfamiliarity with it. The help text included a thorough explanation of the lifetime learning tax credit that telephone interviewers were able to read to respondents unfamiliar with the credit. The other items were accessed only a few times each. The available help text with term definitions was vital in helping telephone interviewers explain unknown terms to respondents. As a result, respondents were able to better understand and answer the survey items.

**Table 4.3. Item-level rates of help text access for B&B:2000/01**

CATI variable name	Label	Number of times help text was accessed	Rate of help text usage <sup>1</sup>	Rate of indeterminacy <sup>2</sup>
D_LIFLNG	Claimed lifetime learning tax credit	286	5.6	10.0
G_FSTPOS	Teaching position held in first job	15	3.6	0.0
G_PRCMP	Complete teacher education program	21	3.2	0.0
D_CREDIT	Claim lifetime learning tax credit next year	112	2.7	0.2
G_EDPGM	Completed a teacher education program	15	1.7	0.0
G_LNFNOW	Taking advantage of loan forgiveness now	4	1.1	0.0
G_EXAM	Taken national/state-level exam	5	1.1	0.0
G_FSTCLS	Sections taught per day—first job	2	1.0	0.0
G_CRTCRS	Taken course for exam—teachers	4	1.0	0.0
G_SUBL2	Long-term substitute—current job	4	1.0	0.0

<sup>1</sup>The rate presented is the number of times the help text for each item was accessed, divided by the number of times that particular item was administered.

<sup>2</sup>The rate of indeterminacy is the number of “don’t know” and “refused” responses, divided by the number of times the item was administered.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

### 4.3 Online Coding

The B&B:2000/01 CATI instrument included tools that allowed computer-assisted online assignment of codes to literal responses for postsecondary education institutions attended, major field of study, occupation, and industry. Online coding systems were designed to improve data quality by capitalizing on the availability of the respondent at the time the coding was performed. Interviewers could request clarification or additional information if a particular text string could not be successfully coded on the first attempt, an advantage not afforded when coding occurs after the interview is complete. Interviewers were trained to use effective probing techniques to ensure that each response was appropriately coded. Because both the literal string and selected code were captured in the data file for major field of study and occupation/industry responses, subsequent quality control recoding by project staff was easily incorporated into data collection procedures.

---

<sup>17</sup>The requirement for item inclusion in the table was a rate of help text usage greater than or equal to 1 percent.

Institutional coding was used to assign six-digit Integrated Postsecondary Educational Data Systems (IPEDS) unit identification numbers for each postsecondary institution that the respondent reported attending. To facilitate coding, the IPEDS coding system asked for the state in which the school was located, followed by the city, and finally the name of the postsecondary institution. The system relied on a look-up table, or coding dictionary, of institutions. This dictionary was constructed from the 1998–99 IPEDS Institutional Characteristics (IC) file. Additional postsecondary information on institutional level and control was retrieved from the look-up table after coding for use for later sections of the interview.

Major field of study, occupation, and industry coding used a dictionary of word and code associations. The online procedures for these coding operations consisted of four steps:

- the interviewer keyed the verbatim text provided by the respondent;
- the dictionary system displayed words that were associated with the words in the text string, and the interviewer was given the choice of either accepting a word that might help in terms of coding, or ignoring a word that was of no help;
- standard descriptors associated with identified codes were displayed for the interviewer; and
- the interviewer selected a standard descriptor that was listed.

The first step in the online coding process after interviews were completed was the upcoding of literal strings that were deemed “uncodable” by telephone interviewers. A string was considered uncodable if the telephone interviewer could not categorize it during the interview. Project staff then attempted to upcode these strings into appropriate categories after the completion of the interview. Table 4.4 summarizes the upcoding effort and indicates that very few strings were uncodable. The elementary/secondary school where teachers were employed and institutional codes (IPEDS) had the highest rate of uncodable items. This was most likely due to two factors: (1) to code uncodable items, it is necessary for there to be an exact match to the school name, so any missing information (such as city or state) made it more difficult to reconcile the problem items; and (2) foreign schools were not included in the IPEDS coding system, making these items uncodable with no possibility of further coding.

**Table 4.4. Success rates for online coding procedures: Upcoding**

<b>Coding procedure</b>	<b>Coding attempts<sup>1</sup></b>	<b>Number originally uncodable</b>	<b>Percent originally uncodable</b>	<b>Percent successfully coded</b>
Elementary/secondary school	1,490	205	13.8	97.1
IPEDS	21,232	1,021	4.8	97.0
Major field of study	8,235	150	1.8	99.4
Occupation	6,912	76	1.1	99.6
Industry	6,307	55	0.9	99.2

<sup>1</sup>Because these items may have been asked multiple times in an interview (e.g., current major, future major), the total number of coding attempts may exceed the total number of completed interviews.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).



The second step to ensure data quality was the recoding process, in which 10 percent of the major field of study, occupation, and industry coding results were then examined and recoded when necessary. Verbatim strings were evaluated for completeness and appropriateness of the assigned codes. Overall, six of the verbatim strings in the sample were too vague to evaluate. Strings for occupation had the highest number of recodes, while industry and major field of study strings required relatively little recoding. Furthermore, none of the recoded cases resulted in a shift across broad categories. Table 4.5 shows the results of the online coding procedures.

**Table 4.5. Success rates for online coding procedures: Recoding**

Coding procedure	Coding attempts sampled	Number too vague	Percent too vague	Number recoded	Percent recoded
Major field of study	778	1	.01	30	3.9
Occupation	642	1	.02	81	12.6
Industry	561	4	.07	15	2.7

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

#### 4.4 CATI Quality Circle Meetings

Quality circle meetings were vital components of the follow-up study operation and evaluation. During these regularly scheduled meetings, interviewers, supervisors, and project technical staff met to discuss issues relevant to locating respondents and conducting CATI interviews in the most effective manner. These meetings proved to be a good tool for communication, as they provided a forum to discuss many elements of the CATI instrument. Telephone interviewers attended the quality circle (QC) meetings on a rotating basis to ensure representation of various experiences, opinions, and challenges faced. Summaries of discussions and decisions were distributed to all telephone interviewers and supervisors in a newsletter. An electronic copy of this newsletter was sent to project staff not in attendance, so those who did not attend the meeting could also benefit. Table 4.6 provides a summary of these meetings.

**Table 4.6. Quality circle meeting summary**

Description of activities	
<b>Number of meetings</b>	Six
<b>Attendees</b>	<ul style="list-style-type: none"> <li>• Project technical staff</li> <li>• Telephone supervisors</li> <li>• Telephone interviewers</li> </ul>
<b>Issues addressed</b>	<ul style="list-style-type: none"> <li>• Clarification of item responses and interpretation of meanings.</li> <li>• Additions to and/or revisions of CATI items and response options.</li> <li>• Changes to help text and procedures.</li> <li>• Submission of problem sheets and notification of supervisors.</li> <li>• Reminders to focus on coding and accuracy.</li> <li>• Stressing the importance of providing detailed case comments.</li> <li>• General morale boosting and reinforcement of positive interviewing techniques.</li> </ul>

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

QC meetings addressed the concerns of project staff regarding the survey instrument and were critical in providing prompt solutions to problems encountered by interviewers. Throughout the duration of the survey, a variety of issues were addressed at the QC meetings. Some of the issues covered in quality circle meetings included the following:

- *Clarification of item responses and interpretation of meanings.* Misinterpretation of questions was addressed consistently. For example, some respondents were claiming to be teachers in areas not considered valid for this study (e.g., college professors, swim instructors). To prevent an interview from proceeding along an inappropriate path, interviewers were instructed to emphasize ‘at the K-12 level’ when asking a respondent if they were currently employed as a teacher, teacher’s aide, or substitute teacher.
- *Changes to the instrument.* Minor modifications to the instrument that were made after interviewer training were explained and demonstrated. This was to ensure that interviewers were aware of the changes and could work with them effectively.
- *Help screens.* Interviewers were reminded of the help text feature, which was available for every CATI item via the F10 function key. The help text screens provided additional explanation to allow interviewers to verify the intent of questions, as well as definitions of unfamiliar terms. Any changes to help text were also discussed. Interviewers were also able to use the F1 function key for quick access to student information, a calculator, roster lines, and case level comments.
- *Problem sheets.* Issues identified on problem sheets and proper documentation procedures were also discussed. Problem sheets were used to convey a variety of information regarding the interview, including data corrections, case anomalies, and areas of confusion.

QC meetings helped to refine interviewer skills and gave project staff feedback that was influential in making the survey clear for respondents and interviewers alike. Interviewers were reminded to focus on coding and accuracy. Overall, the QC meetings were used to help project staff and programmers refine the instrument, to ensure that the most accurate information was obtained during data collection, and to provide reinforcement of positive interviewing techniques.

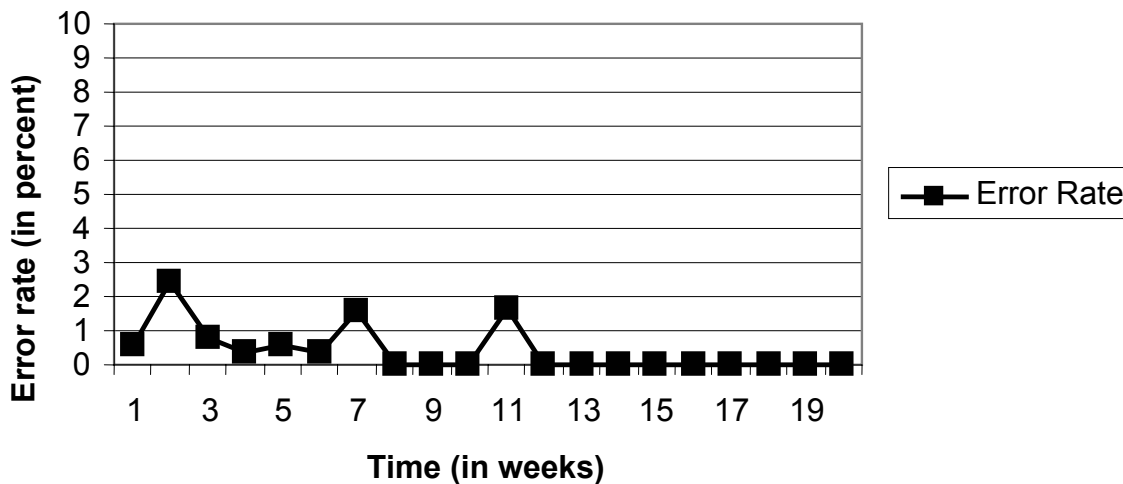
#### **4.5 Quality Assurance CATI Monitoring**

Monitoring of telephone data collection leads to better interviewing and better quality survey data, as well as improvements in costs and efficiency in telephone facilities. Monitoring in the B&B:2000/01 follow-up study helped to meet these important quality objectives:

- reduction in the number of interviewer errors,
- improvement in interviewer performance by reinforcing good interviewer behavior, and
- assessment of the quality of the data being collected.

Monitors listened to up to 20 questions as the interviews were in progress, and for each question evaluated two aspects of the interviewer-respondent interchange: (1) whether the interviewer delivered the question correctly and (2) whether the interviewer keyed the appropriate response. Each of these measures was quantified, and daily, weekly, and cumulative reports were produced for the study's IMS. During the data collection period, 14,952 items were monitored. The majority of the monitoring was conducted during the first half of data collection. Toward the end of data collection, monitoring efforts were scaled back due to the lighter caseload being worked by telephone interviewers, the greater experience of the interviewers, and the satisfaction by project staff that the process was in appropriate control. Figure 4.1 shows error rates for question delivery; figure 4.2 shows error rates for data entry. During data collection, the error rates were monitored to ensure that they were within the upper and lower control limits for these measures.<sup>18</sup>

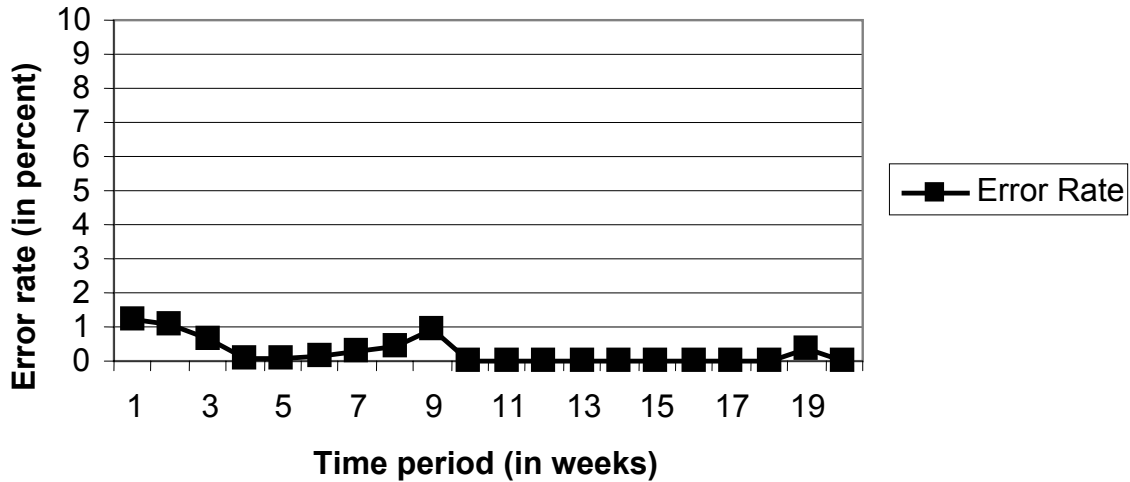
**Figure 4.1. Monitoring error rates for CATI question delivery**



SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

<sup>18</sup> The upper and lower control limits were defined by three times the standard error of the cumulative proportion of errors to the number of questions observed for the period (+3 \* SE for the upper limit; -3 \* SE for the lower limit). These values represent the upper and lower boundaries of expected normal range of statistical variation for the data during the observation period.

Figure 4.2. Monitoring error rates for CATI data entry



SOURCE: U.S. Department of Education, National Center for Education Statistics, 2000–2001 Baccalaureate and Beyond Longitudinal Study (B&B:2000/01).

Throughout the monitoring period, error rates remained within acceptable limits, typically below 1 percent. Among the 14,592 items observed, there were 141 total CATI question delivery errors and 91 total data entry errors. This resulted in overall error rates of less than 1 percent for both question delivery (0.94 percent) and data entry (0.6 percent).