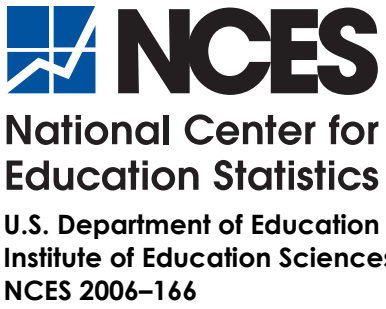


National Center for Education Statistics
U.S. Department of Education
Institute of Education Sciences
NCES 2006–166

1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03)

Methodology Report

November 2005

This page left intentionally blank.

National Center for Education Statistics
U.S. Department of Education
Institute of Education Sciences
NCES 2006–166

1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03)

Methodology Report

November 2005

Jennifer S. Wine
Melissa B. Cominole
Sara Wheeless
Kristin Dudley
Jeff Franklin
RTI International*

Kristin Perry
Project Officer
National Center for Education Statistics

* RTI International is a trade name of Research Triangle Institute.

U.S. Department of Education
Margaret Spellings
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Education Statistics
Mark Schneider
Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high-priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high-quality data to the U.S. Department of Education, Congress, state and other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street, NW
Washington, DC 20006-5651

November 2005

The NCES World Wide Web Home Page address is http://nces.ed.gov/
The NCES World Wide Web Electronic Catalog is http://nces.ed.gov/pubsearch/

Suggested Citation
Wine, J.S., Cominole, M.B., Wheeless, S., Dudley, K., and Franklin, J. (2005). *1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03) Methodology Report* (NCES 2006–166). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Contact:

Aurora D'Amico
Aurora.D'Amico@ed.gov
(202) 502-7334

Executive Summary

Introduction

The 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03), sponsored by the National Center for Education Statistics (NCES), U.S. Department of Education, followed a cohort of students who earned bachelor's degrees during the 1992–93 academic year. These students were first interviewed in 1993, as part of the 1993 National Postsecondary Student Aid Study (NPSAS:93). A year later, a follow-up interview was conducted (B&B:93/94) and transcripts were collected from students' undergraduate institutions and coded, and in 1997, a second follow-up interview was conducted (B&B:93/97). B&B:93/03 is the third and final follow-up interview with the class of 1993. This interview, which took place in 2003, focused on postbaccalaureate education, employment and career development, family formation, and finances. The B&B:93 series of interviews has also included specific sections focused on those bachelor's degree recipients who trained to be elementary/secondary teachers (kindergarten through grade 12), and those who have entered or plan to enter teaching, in the first 10 years following bachelor's degree award. This report describes the procedures and results of the full-scale implementation of the 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

Sampling Design

The target population for the B&B:93 study consisted of those individuals who were eligible to participate in NPSAS:93 and were awarded the bachelor's degree by a postsecondary institution in the United States, the District of Columbia, or Puerto Rico. The B&B:93 cohort consisted of both students who completed the NPSAS:93 interview and were identified to be baccalaureate recipients and those NPSAS:93 nonrespondents who were potentially eligible for B&B who had at least some data (from either the institutional records or interviewing). Following NPSAS:93, 16,320[**] baccalaureate degree recipients were identified for participation in the first follow-up interview. By the second follow-up interview, B&B:93/97, 11,190 cases were retained for participation: 10,080 computer-assisted telephone interviewing (CATI)-eligible cases, 1,090 transcript-eligible cases, and 20 cases for which eligibility was unknown for both components. All 10,090 B&B:93/97 respondents were included for participation in B&B:93/03. In addition, a subsample of 360 B&B:93/97 nonrespondents was also included. After removing cases identified as deceased, the starting sample for B&B:93/03 was 10,440.

Instrumentation

For the first time, students were offered the opportunity to conduct their own B&B interview via the Internet. A single, web-based interview was designed and programmed for use as a self-administered interview, a telephone interview, and an in-person interview. In addition,

[*] The numbers appearing in the tables and text of this report have been rounded to the nearest tens to maintain the confidentiality of study participants.

a website was developed to launch the self-administered interview, to provide additional study information, and to collect updated student locating information.

The interview focused on students' activities in the 6 years since the last follow-up interview, B&B:93/97. The first section collected information on additional education pursued since 1997 whether through a formal graduate or undergraduate program, as part of occupational certification/licensure or employment training, or for personal enrichment. The next section focused on current employment, employment patterns, and career development. A separate section addressed employment patterns and job satisfaction for new teachers and those who have been a part of the teacher pipeline throughout the 10 years since degree award. The next section collected information on students' marital status, family, civic involvement, and disability status. The final section addressed finances, including education loan debt, assets, and income.

Data Collection Design and Outcomes

Training

Separate training programs were developed for those staff monitoring the Help Desk, made available to sample members completing the self-administered interview and to those interviewers conducting telephone and in-person interviews. Training topics included an overview of B&B:93/03, case management, quality control, solutions to common problems encountered using the self-administered interview, ways to establish effective relationships with sample members and other contacts, the nature of the data to be collected, and the organization and operation of the web-based interview. Tracing specialists received an abbreviated training specific to the needs of locating B&B:93/03 sample members.

Interviewing

The self-administered interview was made available to sample members beginning in February 2003. After 3 weeks, telephone interviewing began with those sample members who had not yet completed the self-administered interview. About 3 months following the start of telephone interviewing, field interviewers began tracing and interviewing nonrespondents whose last known address was in one of 30 geographic clusters.

From the starting sample of 10,440 about 40 individuals were found to be deceased and another 10 were determined to be study ineligible. The unweighted locate rate among the remaining sample members was 93 percent. Of those located, 92 percent completed the interview for an overall unweighted response rate of 86 percent. Among respondents, 38 percent completed the self-administered interview on the Internet, 57 percent completed a telephone interview, and the remaining 5 percent were interviewed in person.

Use of Incentives

Incentives were offered to sample members at two different points during data collection. First, sample members were offered a $20 cash incentive for completing the self-administered interview within the first 3 weeks of data collection, prior to the start of telephone interviewing. Of those who completed the self-administered interview, 47 percent did so during the incentive

period. Additionally, an incentive was used to reduce nonresponse among four groups: those who refused to be interviewed, those who could not be reached by telephone, those for whom only a contact person could be reached, and those who started but did not finish the self-administered interview. Overall, 55 percent of sample members falling into one of the four groups completed the interview following the offer of a nonresponse incentive.

Refusal Conversion

The ability of interviewers to gain the cooperation of sample members was important to the success of B&B:93/03. Among telephone interviewers was a group of refusal conversion specialists trained in converting sample members who have refused to complete the interview. From the point when a sample member refused, the case was handled only by these conversion specialists. In B&B:93/03, slightly less than 10 percent of sample members ever refused to participate in the interview. Of those, 49 percent completed the interview.

Interview Burden

Time in the B&B:93/03 interview was calculated separately according to whether the interview was self-administered or interviewer-administered. Self-administered interviews averaged almost 37 minutes, of which almost 11 minutes was due to the time required to transmit data to and from the respondent. Transit times varied considerably depending on the type of internet connection used. Interviewer-administered interviews, both telephone and in-person, averaged 35 minutes, with about 6 minutes required to transmit data (telephone interviews only). While the overall time to complete the interview was longer for self-administered respondents, the actual interview time was shorter for self-administered respondents.

Usability of the Instrument

Help text was available for every screen of the B&B:93/03 instrument. Help text screens displayed instructions on how to enter responses, the type of information requested, and definitions of words or phrases within an item. Help text usage rates were consistently under 1 percent.

The B&B:93/03 instrument also included tools that allowed online coding of literal responses for occupation, industry, major/field of study, and area of licensure/certification. Throughout data collection, coding experts examined samples of each set of coding results for completeness and the correctness of codes selected by self-administered interview respondents and interviewers. A comparison of recode results by mode of data collection showed that interviewers tended to do somewhat better than sample members in selecting the correct code.

Indeterminate Responses

To minimize item-level nonresponse, the B&B:93/03 interview was designed without explicit “don’t know” and “refuse to answer” options on screen. Instead, respondents could use the “continue” button to proceed without answering an item. Pop-up boxes and response conversion text were used to encourage respondents to provide an answer. As a result, only 20 of over 650 full-scale interview items had missing data at a rate of 10 percent or more.

Analysis Weights

Cross-sectional weights were developed for analyzing respondents to the B&B:93/03 interview. In addition, a longitudinal weight was constructed for analyzing students who participated in all four interviews—NPSAS:93, B&B:93/94, B&B:93/97, and B&B:93/03. Variances were computed using the Taylor series and balanced repeated replications (BRR) techniques.

Data Files

The B&B:93/03 study was the fourth of four interviews with the B&B:93/03 cohort. The dataset, therefore, includes the derived variable and interview files for all four interviews. Also included are data collected from transcript coding, institution records, government databases, and admission test vendors throughout the period from the NPSAS:93 interview through the B&B:93/03 interview.

Products

In addition to the methodology report, NCES plans to release the following major products for B&B:93/03: a public use Data Analysis System (DAS), restricted use research files with an associated electronic codebook (ECB), an E.D. Tabulation of general findings, a descriptive summary of significant findings focusing on outcomes for bachelor's degree recipients in the 10 years since degree award, and a descriptive report on the current status of elementary/secondary teachers who began teaching at some time in the 10-year period since degree award.

Foreword

This report describes the methods and procedures used for the full-scale data collection effort of the 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03). These students, who earned bachelor's degrees during the 1992–93 academic year, were first interviewed in 1993, as part of the 1993 National Postsecondary Student Aid Study (NPSAS:93). They were subsequently interviewed and undergraduate transcripts were collected during the first follow-up study in 1994 (B&B:93/94). Three years later, the second follow-up interview was conducted (B&B:93/97). B&B:93/03 is the third and final follow-up interview of the B&B:93 cohort, 10 years following completion of the bachelor's degree. While, like all longitudinal studies, sample maintenance has presented a significant challenge over the 10-year span of the B&B studies, B&B:93/03 offers a unique opportunity to assess the value of the baccalaureate degree to both the individual and to society, at a time when sample members are most likely to be established in their careers.

For the first time, B&B:93/03 offered students the opportunity to conduct their own interview over the Internet. A single, web-based instrument was designed and programmed for use as a self-administered interview, a telephone interview, and an in-person interview. A study website was also made available for launching the self-administered interview as well as for providing information about the study, contact information for project staff, and an address update capability.

Evaluation of the procedures used in the full-scale study were developed and refined as part of the field test data collection conducted in 2002. We hope that the information provided here will be useful to a wide range of interested readers and that the results reported in the forthcoming descriptive summary report and teacher pipeline report will encourage others to use the Baccalaureate and Beyond Longitudinal Study data.

C. Dennis Carroll
Associate Commissioner
Postsecondary Studies Division

This page left intentionally blank.

Acknowledgments

The authors gratefully acknowledge the assistance of staff members of the National Center for Education Statistics (NCES) and the Institute of Education Sciences (IES) for their advice, guidance, and review in conducting the 1993/03 Baccalaureate and Beyond Longitudinal Study and in preparing this document. We are particularly grateful to C. Dennis Carroll, Associate Commissioner, Postsecondary Studies Division; James Griffith, Program Director for Postsecondary Longitudinal Studies and Sample Surveys; Paula Knepper, Senior Technical Advisor for Postsecondary Studies; Tracy Hunt-White, Statistician, Postsecondary Studies Division; and Kristin Perry, the Baccalaureate and Beyond Longitudinal Study Project Officer, who reviewed the report at various stages and provided useful feedback. We also wish to thank the IES/NCES reviewers, who read earlier drafts of this report and offered many helpful suggestions.

Particular thanks are extended as well to the study Technical Review Panel members who provided considerable insight and guidance in the development of the design and instrumentation for this study. We also extend our thanks to the project staff members of the two contractors, RTI International* and MPR Associates. A number of staff from these organizations—including statisticians, analysts, survey managers, programmers, data collectors, and interviewers—too numerous to name here, worked long hours on this study. At RTI, we are especially indebted to our editor, Sallie Fiore, and our document preparation specialists Lynne Hawley and Sharon Powell who produced the drafts and final versions of this report.

Most of all, we are greatly indebted to the many postsecondary education institutions, students, former students, and their parents, relatives, and friends, who unselfishly gave of their time to provide study data and locating information.

* RTI International is a trade name of Research Triangle Institute.

This page left intentionally blank.

Table of Contents

List of Tables

List of Figures

This page left intentionally blank.

Chapter 1
Overview of B&B:93/03

This document describes the procedures and results of the full-scale implementation of the 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03). RTI International (RTI), with the assistance of MPR Associates, Inc. (MPR), conducted the study for the National Center for Education Statistics (NCES) of the U.S. Department of Education (Contract No. ED-01-CO-0098), as authorized under Section 404(a) of the National Statistics Act of 1994, 20 U.S.C. 9001 et seq. (2002).

This introductory chapter provides an overview of the background, purposes, schedule, and products of the B&B:93/03 study. The second chapter describes the design and methods used during the study. Outcomes of data collection are presented in chapter 3. Evaluations of the quality of data collected are provided in chapter 4, and a description of the procedures used to create B&B:93/03 data files is presented in chapter 5. Sample weighting and variance estimation are discussed in chapter 6. Materials used during the field test are provided as appendices to the report and cited, where appropriate, in the text. The design and results of the field test study are presented elsewhere (Wine et al. 2004).

1.1 Background and Objectives of B&B

The Baccalaureate and Beyond Longitudinal Study (B&B), which focuses on the experiences of baccalaureate degree recipients over time, was designed to collect important policy-relevant information regarding the occupational and further educational outcomes of earning a baccalaureate degree. The major issues B&B addresses concern the relative value to the individual and to society of earning the bachelor's degree. Specific topics of interest include the length of time to complete the bachelor's degree, the pursuit of additional postsecondary education beyond the bachelor's degree, and employment outcomes. Another focus of B&B has been the teacher pipeline, that is, the progression of those who pursue teaching as a career as well as those who have taken steps to prepare for a career in teaching.

Potential sample members were first identified and selected to participate in B&B:93/03 as part of the cross-sectional 1992–93 National Postsecondary Student Aid Study (NPSAS:93) of postsecondary undergraduate, graduate, and professional students. NPSAS:93 focused on how students and their families financed their postsecondary education. To be eligible to participate in the B&B:93 longitudinal study, sample members had to have been eligible for NPSAS:93 and earned a bachelor's degree during the 1992–93 school year. The NPSAS:93 base-year interview collected information from students, institutions, and parents on background characteristics, enrollment, employment, and education financing, including financial aid. Students being awarded the bachelor's degree at the time of the NPSAS:93 interview were eligible for membership in the B&B:93 longitudinal cohort and, as part of their NPSAS:93 interview, were asked additional questions about their plans for the future, particularly graduate education and any plans to pursue a career in teaching kindergarten through grade 12 (K–12).

The B&B:93 cohort was interviewed again 1 year following degree completion (B&B:93/94). The interview itself covered a number of topics related to education since receipt of the bachelor's degree, including job search and the transition into employment, job training, family formation, civic participation, and finances (including income, student loans, and other debt). In addition, undergraduate transcripts were collected from the schools from which B&B:93 sample members earned their bachelor's degrees. As part of the transcript data collection, school-level information (e.g., information from course catalogs and grading systems) was collected for each sample school. Student-level data, such as major and minor fields of study, grade point average, courses taken, and grades earned, were coded for each student within a school. Transcripts from transfer schools were also coded, when available.

A second follow-up interview with the B&B:93 cohort was conducted in 1997, 4 years following bachelor's degree completion (B&B:93/97). This second follow-up interview collected detailed information on postbaccalaureate enrollment, including degrees sought, enrollment intensity and duration, finances, and degree attainment. Employment information and experiences, such as the number of jobs held since the last interview, occupations, salaries, benefits, and job satisfaction, were also collected. Those already in or newly identified for teaching careers were asked questions about their preparation to teach, work experience at the K–12 level, and satisfaction with teaching as a career. In addition to questions about education and employment, the 1997 interview continued to update information on family formation and civic participation.

The final follow-up interview of the B&B:93 cohort in 2003 (B&B:93/03), conducted 10 years following degree completion, allowed further study of the issues already addressed by the preceding follow-up studies. The 2003 interview covered topics related to continuing education, degree attainment, employment, career choice, family formation, and finances. Additionally, respondents were asked to reflect on the value of their undergraduate education and any other education obtained since receiving the bachelor's degree to their lives now. It also contained a separate set of questions directed at new entrants to the teacher pipeline, as well as those who were continuing in or left teaching since the last interview.

The remainder of this report provides details on the B&B:93/03 sampling design and data collection outcomes. It also presents the results of analyses conducted to evaluate the quality of the B&B:93/03 data, as well as an overview of the B&B:93/03 data files and variables. Finally, the report provides a discussion of procedures and results related to weighting, variance estimation, precision, nonresponse bias, and imputations. Unless otherwise indicated, a criterion probability level of 0.05 was used for all tests of significance.

1.2 Schedule and Products of B&B:93/03

The B&B:93 full-scale study was preceded by a field test. Full-scale data collection took place between February and September 2003. Dates of other key activities are also presented in the operational schedule shown in table 1.

Table 1. Operational schedule for B&B:93/03

Activity	Start date	End date
Field test		
Sampling	December 2001	December 2001
Tracing	September 2001	June 2002
Web/self-administered interviewing	April 2002	July 2002
Telephone interviewing	April 2002	July 2002
Field data collection	June 2002	July 2002
Data files and documentation	March 2002	September 2002
Field test report	June 2002	June 2003
Full-scale study		
Sampling	December 2002	December 2002
Tracing	October 2002	September 2003
Web/self-administered interviewing	February 2003	September 2003
Telephone interviewing	March 2003	September 2003
Field data collection	June 2003	September 2003
Data files and documentation	February 2003	June 2005
Methodology report	March 2003	June 2005
Data Analysis System (DAS)	June 2003	June 2005
Descriptive survey report	October 2003	June 2005
Teacher report	October 2003	June 2005

SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

The major products of B&B:93/03 include the following:

- A bibliography of publications using data for the B&B:93 cohort;
- Methodology reports (one each for the field test and the full-scale study) that describe all aspects of the data collection effort;
- Restricted-use data files and documentation for research data users;
- A Data Analysis System for public access to the B&B:93 longitudinal data, including the base-year interview, three follow-up interviews, and transcript abstraction;
- Special tabulations of issues of interest to the higher education community, as determined by NCES; and
- A descriptive overview report for the B&B:93/03 data collection. This report will present significant findings across a broad spectrum of outcomes for bachelor's degree recipients 10 years later. Major milestones such as graduate degree attainment, employment status, family roles, community service, and reflections on undergraduate education will be described, exploring differences by demographic characteristics and undergraduate experiences.

- A report on K-12 teachers. This report will describe the teaching experiences and preparation of 1992–93 bachelor's degree recipients in the 10 years following their college completion.

Chapter 2
Design and Method

This chapter describes the sampling design used for the Baccalaureate and Beyond Longitudinal Study (B&B), beginning with an overview of the sampling procedures used since the base-year study (1993 National Postsecondary Student Aid Study [NPSAS:93]). Sampling procedures and data collection design for the final follow-up study (B&B:93/03) are presented in detail, including sample member locating and contacting activities and interview design. The many systems supporting B&B:93/03 are also described.

2.1 Sampling Design

The target population for the B&B:93 full-scale study consisted of those individuals who were eligible to participate in NPSAS:93 and were awarded the bachelor's degree by a postsecondary institution in the United States or Puerto Rico. Members of the B&B:93 cohort were identified during NPSAS:93, which served as the base year for the longitudinal study. The B&B:93 cohort consisted of both students who completed the NPSAS:93 interview and were identified to be baccalaureate recipients, and those NPSAS:93 nonrespondents who were potentially eligible for B&B who had at least some data (from either the institutional records or the computer-assisted telephone interviewing [CATI]).[1] The NPSAS:93 sampling design was a two-stage design in which eligible institutions were selected first, and then eligible students were selected from eligible, participating institutions. The sampling procedures used to select institutions and students in the base-year and follow-up studies are described below. The numbers shown in the text and tables of this report have been rounded to the nearest tens and hundreds to maintain the confidentiality of study participants.

2.1.1 NPSAS:93 Institution Universe

The institution universe for the B&B was the set of institutions eligible for NPSAS:93. To be eligible for NPSAS:93, an institution had to do the following during the 1992–93 academic year:

- offer an education program designed for persons who have completed secondary education;
- offer an academically, occupationally, or vocationally oriented program of study;
- offer courses to students not employed by the institution;
- offer more than just correspondence courses;
- offer at least one program requiring at least 3 months or 300 clock hours of instruction;

[1] NPSAS:93 nonrespondents were identified as potentially eligible for inclusion in the B&B cohort if there was information from institutional records indicating that they had received, or expected to receive, a baccalaureate degree between July 1, 1992 and June 30, 1993.

- be located in one of the 50 states, the District of Columbia, or Puerto Rico; and
- not be a U.S. service academy.

U.S. service academies were excluded from participation because of their atypical funding and tuition base. Also ineligible were institutions offering only avocational, recreational, remedial, or correspondence courses; institutions offering only noncredit continuing education units (CEUs); schools whose only purpose was to prepare students to take a particular examination (e.g., the Certified Public Accountant [CPA] or Bar exams); institutions offering only programs of study which required less than 3 months or 300 contact hours of instruction; and branch campuses of U.S. institutions in foreign countries.

2.1.2 NPSAS:93 Institution Sample

The institution-level sampling frame for NPSAS:93 was constructed from the 1990–91 Integrated Postsecondary Education Data System (IPEDS) Institutional Characteristics (IC) file. Excluded from the frame were those NPSAS:93-ineligible institutions described above and five institutions determined to be ineligible based on discrepancies in the IPEDS enrollment data. The resulting sampling frame contained 10,140 institutions that appeared to be eligible for NPSAS:93 based on their 1990–91 IPEDS IC data.

Geographic areas defined by three-digit postal zip codes were used as the basis for creating primary sampling units (PSUs) of nearly equal sizes to ensure statistical efficiency. The final area sampling frame contained 291 PSUs, of which 86 were certainty and the remaining 205 were noncertainty.[2] The final NPSAS:93 sample contained the 86 certainty PSUs and a sample of 90 of the 205 noncertainty PSUs selected with probabilities proportional to size, for a total of 176 PSUs. All institutions within the 176 sample PSUs were then combined into a single frame, stratified by the 22 strata shown in table 2. Sample institutions were selected using measures of size that were proportional to the expected sample allocation for the institution.

Although the IPEDS frame provided good coverage of the population of postsecondary institutions, a supplemental sample was selected from the Office of Postsecondary Education's Institutional Data System (OPE-IDS) file of institutions participating in the Pell Grant and Stafford Loan Programs as of April 15, 1992, to improve coverage. The OPE-IDS file was subset to those institutions located in the 176 survey PSUs based on zip codes.

Within the 22 institutional strata for the primary sample of institutions, the sample was implicitly stratified by sorting the frame units by the following variables:

- Office of Business Economics (OBE) region;[3]
- state;
- PSU; and

[2] PSUs that contained the largest institutions were defined to be certainty PSUs. For more information, see *Methodology Report for the National Postsecondary Student Aid Study, 1992-93* (Loft et al. 1995).
[3] The OBE region classifications are as follows: 1 includes CT, ME, MA, NH, RI, VT; 2 includes DE, DC, MD NJ, NY, PA; 3 includes IL, IN, MI, OH, WI; 4 includes IA, KS, MN, MO, NE, ND, SD; 5 includes AL, AR, FL, GA, KY, LA, MS, NC, PR, SC, TN, VA, WV; 6 includes AZ, NM, OK, TX; 7 includes CO, ID, MT, UT, WY; 8 includes CA, HI, NV, OR, WA; and 9 includes AK and HI.

- measure of size.

A sample of 1,360 institutions (720 from the certainty PSUs and 640 from the noncertainty PSUs) was selected for the primary sample from the IPEDS frame.

The supplemental sample from OPE-IDS file was implicitly stratified by sorting by the following variables:

- level (less-than-2-year, 2-year, or 4-year);
- control (public, private not-for-profit, or private for-profit); and
- the Study ID from the OPE-IDS file (in order to produce a unique frame ordering).

Twenty-two institutions were selected from the OPE-IDS frame to create the supplemental sample.

The overall institution sample sizes are shown in table 2 for each of the 22 institutional sampling strata. The expected frequency of selection exceeded unity (1.00) for some institutions, and those institutions were included in the sample with certainty. The numbers of certainty and noncertainty selections of institutions for the primary sample are shown for each stratum in table 3.

Table 2. 1992–93 National Postsecondary Student Aid Study (NPSAS:93) institutional sample sizes, by institutional stratum

Institutional stratum	Total institutional sample[1]	86 Certainty PSUs			90 Noncertainty PSUs		
		Total number of institutions from certainty PSUs	Number of institutions from IPEDS[2]	Number of institutions from OPE-IDS[3]	Total number of institutions from non-certainty PSUs	Number of institutions from IPEDS[2]	Number of institutions from OPE-IDS[3]
Total	**1,390**	**730**	**720**	**10**	**660**	**640**	**10**
1. Public 4-year first-professional high education [4]	20	10	10	#	10	10	#
2. Public 4-year first-professional low education	100	80	80	#	20	20	#
3. Private not-for-profit 4-year first-professional high education [5]	80	50	50	#	30	30	#
4. Private not-for-profit 4-year first-professional low education	80	50	50	#	30	30	#
5. Public 4-year doctor's high education [4]	10	10	10	#	#	#	#
6. Public 4-year doctor's low education	40	20	20	#	20	20	#
7. Private not-for-profit 4-year doctor's high education [5]	20	10	10	#	10	10	#
8. Private not-for-profit 4-year doctor's low education	20	10	10	#	10	10	#
9. Public 4-year master's high education [6]	30	10	10	#	20	20	#
10. Public 4-year master's low education	120	50	50	#	70	70	#
11. Private not-for-profit 4-year master's high education [6]	10	#	#	#	10	10	#
12. Private not-for-profit 4-year master's low education	130	60	60	#	60	60	#
13. Public 4-year bachelor's high education [6]	10	#	#	#	10	10	#
14. Public 4-year bachelor's low education	40	10	10	#	20	20	#
15. Private not-for-profit 4-year bachelor's high education [6]	10	#	#	#	10	10	#
16. Private not-for-profit 4-year bachelor's low education	80	30	30	#	50	50	#
17. Public 2-year	220	100	100	#	120	110	#
18. Private not-for-profit 2-year	20	10	10	#	10	10	#
19. Private for-profit 2-year	50	30	30	#	20	20	#
20. Public less-than-2-year	50	30	30	#	30	20	#
21. Private not-for-profit less-than-2-year	50	30	30	#	20	20	#
22. Private for-profit less-than-2-year	220	110	110	#	110	100	#

Rounds to zero.
[1] This total includes institutions from both the Integrated Postsecondary Education Data System (IPEDS) file and the Office of Postsecondary Education's Institutional Data System (OPE-IDS) file.
[2] Primary sample.
[3] Supplemental sample.
[4] More than 15 percent of baccalaureate degrees were awarded in education.
[5] Any baccalaureate degrees awarded in education.
[6] More than 25 percent of baccalaureate degrees were awarded in education.
NOTE: PSU = primary sampling unit. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1992–93 National Postsecondary Student Aid Study (NPSAS:93).

Table 3. Number of certainty and noncertainty institutions in the 1992–93 National Postsecondary Student Aid Study (NPSAS:93) primary institutional sample, by institutional stratum

Institutional stratum	Total[1]	86 Certainty PSUs		90 Noncertainty PSUs	
		Certainty institutions	Non-certainty institutions	Certainty institutions	Non-certainty institutions
Total	1,360	300	420	290	360
1. Public 4-year first-professional high education[2]	20	10	#	10	#
2. Public 4-year first-professional low education	100	80	#	#	10
3. Private not-for-profit 4-year first-professional high education[3]	80	40	10	30	#
4. Private not-for-profit 4-year first-professional low education	80	20	40	10	20
5. Public 4-year doctor's high education[2]	10	10	#	#	#
6. Public 4-year doctor's low education	40	20	#	20	#
7. Private not-for-profit 4-year doctor's high education[3]	20	10	10	10	#
8. Private not-for-profit 4-year doctor's low education	20	#	10	#	10
9. Public 4-year master's high education[4]	30	10	#	20	#
10. Public 4-year master's low education	120	50	#	70	#
11. Private not-for-profit 4-year master's high education[4]	10	#	#	#	10
12. Private not-for-profit 4-year master's low education	130	10	50	30	40
13. Public 4-year bachelor's high education[4]	10	#	#	10	#
14. Public 4-year bachelor's low education	40	#	10	20	#
15. Private not-for-profit 4-year bachelor's high education[4]	10	#	#	#	10
16. Private not-for-profit 4-year bachelor's low education	80	#	30	10	40
17. Public 2-year	210	10	90	30	80
18. Private not-for-profit 2-year	20	#	10	#	10
19. Private for-profit 2-year	50	#	20	#	20
20. Public less-than-2-year	50	10	20	#	10
21. Private not-for-profit less-than-2-year	40	10	20	10	10
22. Private for-profit less-than-2-year	210	10	100	10	90

Rounds to zero.
[1] Numbers presented here are based only on the primary sample of institutions, i.e., those selected from the Integrated Postsecondary Education Data System (IPEDS) frame. Institutions selected from the Office of Postsecondary Education's Institutional Data System (OPE-IDS) frame are not presented in this table.
[2] More than 15 percent of baccalaureate degrees were awarded in education.
[3] Any baccalaureate degrees awarded in education.
[4] More than 25 percent of baccalaureate degrees were awarded in education.
NOTE: PSU = primary sampling unit. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1992–93 National Postsecondary Student Aid Study (NPSAS:93).

2.1.3 NPSAS:93 Student Universe

Students eligible for inclusion in the B&B:93 longitudinal cohort were those students who were eligible for NPSAS:93. NPSAS-eligible students had to have been enrolled in a NPSAS-eligible institution between July 1, 1992, and June 30, 1993, and

- enrolled in *either* (a) course(s) for credit toward a degree or formal award; (b) a degree or formal award program of at least 3 months; or (c) an academically, occupationally, or vocationally specific program requiring at least 3 months or 300 clock hours of instruction;
- not currently enrolled in high school; and
- not currently *solely* in a general equivalency diploma (GED) or other high school completion program.

Additionally, to be eligible for the B&B:93 cohort, students could have been awarded a baccalaureate degree at some time between July 1, 1992, and June 30, 1993, irrespective of their enrollment status during the academic year. The sampling process is described in more detail below.

2.1.4 NPSAS:93 Student Sample

To create the NPSAS student sampling frame, each sample institution was asked to provide a list of all students enrolled during the NPSAS year (July 1, 1992–June 30, 1993) and those awarded a baccalaureate degree at some point during that year, according to eligibility criteria provided to the institutions. Stratified systematic sampling was used to facilitate sampling from lists. For each sample institution, student sampling rates were determined for each of five student sampling strata:

- business major baccalaureates;
- other baccalaureate recipients;
- other undergraduates, including enrollees at less-than-4-year institutions;
- graduate students; and
- first-professional students.

The sampling rates depended on the overall population sampling rates for the five types of students, the probability of selecting the institution, and a requirement for a minimum of 40 sample students per institution whenever possible.

Sample institutions identified those students eligible to receive the bachelor's degree during the 1992–93 academic year for inclusion in the B&B:93 cohort. In addition, during the CATI, those students who indicated having received a baccalaureate degree during the 1992–93 academic year were also included. From the NPSAS:93 sample, 16,320 baccalaureate degree recipients were identified for participation in the B&B:93 cohort.

2.1.5 B&B:93 Cohort—Follow-up Samples

Sampling procedures used to maintain the B&B:93 cohort through the follow-up studies are summarized below. Greater detail of the sampling procedures used in each follow-up can be found in the respective methodology reports (Loft et al. 1995; Green et al. 1996; Green et al. 1999). Table 4 presents the distributions of the student samples from each wave.

2.1.5.1 First follow-up—B&B:93/94

As discussed above, 16,320 baccalaureate degree recipients were identified for inclusion in the B&B:93 cohort from institutionally-provided lists of students who were eligible for graduation or who indicated having graduated in the 1992–93 academic year during the CATI interview. All 11,810 of the identified students who completed the NPSAS:93 interview were retained for the B&B:93 cohort. Also retained were 370 student nonrespondents for whom NPSAS parent data were available that indicated that the student received the bachelor's degree during 1992–93. Additionally, a 10 percent subsample of the remaining eligible cases with at least some data was included, for a total of 12,730[4] eligible cases. It became apparent during data collection that many of the nonrespondents and potentially eligible cases were actually ineligible. Because of the costs associated with the ineligible students, only a subsample of the nonrespondents and potentially eligible students was selected, reducing the B&B sample size to 12,480 (see table 4).

[4] The numbers in the tables and text of this report have been rounded to the nearest tens to maintain the confidentiality of study participants.

Table 4. Summary of the B&B:93 cohort sample: 2003

Interviews with B&B:93 cohort	Sample size
Total	10,440
Original B&B cohort from NPSAS:93	
Student respondents	11,810
Student nonrespondents with parent data	370[1]
Subsample of other student nonrespondents	160[1]
Additional B&B cases identified during data processing (NPSAS student respondents identified as potentially eligible for B&B)	200[1]
Number retained for B&B:93/94	12,480
Respondents in B&B:93/94	10,080
Cases defined as eligible based on transcript study [2]	1,090
Eligibility unknown in B&B:93/94	20
Ineligible in B&B:93/94	-1,290
Number retained for B&B:93/97	11,190
Respondents in B&B:93/97	10,090
Nonrespondents in B&B:93/97	1,070
Ineligible in B&B:93/97	-30
Number retained for B&B:93/03	11,160
Cases identified as deceased during B&B:93/03 advance tracing death search	-10[3]
B&B:93/97 respondents	10,090
Subsample of B&B:93/97 nonrespondents[4]	360
B&B:93/97 nonrespondents not subsampled	710

[1] Many of the students in these groups were determined to be ineligible for B&B. As a result, the groups were subsampled.
[2] When available, transcript data were used to assist in B&B eligibility determination. See the U.S. Department of Education, National Center for Education Statistics. *Baccalaureate and Beyond Longitudinal Study: 1993/94 First Follow-up Methodology Report*, NCES 96-149, by Patricia J. Green, Sharon L. Meyers, Pamela Giese, Joan Law, Howard M. Speizer, and Vicki Staebler Tardino. Project Officer, Paula Knepper. Washington, DC: 1996 for a description of the transcript component of the B&B:93/94 study.
[3] Among the deceased were both B&B:93/97 respondents and nonrespondents.
[4] A subsample of about one-third of the B&B:93/97 nonrespondents was included in the B&B:93/03 sample.
NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

2.1.5.2 Second follow-up—B&B:93/97

B&B:93/94 included a transcript component in which eligibility for the B&B:93 cohort was determined for base-year nonrespondents. After data collection was complete for the first follow-up, additional ineligible cases were found in the cohort based on the information obtained from the transcript data. Sample members were retained for follow-up in later rounds if they were found to be eligible in either the CATI or the transcript component. In total, 11,190 cases were retained for the second follow-up, B&B:93/97: 10,080 CATI-eligible cases, 1,090 transcript-eligible cases, and 20 cases for which eligibility was unknown for both components.

2.1.5.3 Third follow-up—B&B:93/03

All of the B&B:93/97 respondents were included in the B&B:93/03 sample. However, because it is more difficult and expensive to locate and interview prior nonrespondents, a subsample of about one-third, or 360 of B&B:93/97 nonrespondents was included. After removing cases identified as deceased, the final sample for B&B:93/03 was 10,440 (see table 4).

The subsample of nonrespondents for B&B:93/03 was stratified by respondent characteristics to ensure that a sufficient range of respondent characteristics was represented. Specifically, controlling for type of institution, age of student, and whether the student was in the teaching pipeline was important for ensuring representativeness of the sample. In addition, stratification by advance tracing outcome and prior-round response status was used to oversample students who were most likely to be located and interviewed.

Three nonrespondent subsampling strata were defined in terms of the B&B:93/03 advance tracing status:

- Students who were located during the advance tracing process (these students are most likely to be found and interviewed);
- Students who were not located during the advance tracing process, but for whom another contact was located; and
- Students who were not located, and for whom no other contact was located (these students are least likely to be found and interviewed).

These three sampling strata were further subdivided based on the control of the base-year institutions and on the response status to B&B:93/94 (respondent or nonrespondent). The final sampling strata for selecting the subsample of B&B:93/97 nonrespondents were defined as shown in table 5.

An indicator of whether the student is in the teaching pipeline and the age of the student were also considered as part of the stratification. However, the "teacher pipeline" indicator was not available for the 460 students who were nonrespondents to both the 1993 and 1997 survey. Of the remainder, almost all were classified as not in the pipeline. Further division based on student age would have created some strata with very few eligible students. For these reasons, age and teacher pipeline status were not used in the stratification. Instead, the file was sorted by age within subsampling strata prior to selecting the sample.

After the nonrespondent sampling strata were defined, a sample allocation was developed that minimized the overall relative cost, subject to constraints that limited the unequal weighting effect, both overall and for each of the 16 B&B strata.[5] The sampling rates and the resulting sample allocations are shown in table 5.

A stratified sample of 360 nonrespondents was selected with probabilities proportional to their B&B:93/94 base weights[6] using the sample allocation shown in table 5. Selection of the

[5] The B&B strata are the first 16 of the 22 NPSAS:93 institutional strata shown in table 2 that correspond to 4-year institutions.
[6] The weight used was the B&B:93/94 base weight (BNBWT0).

B&B:93/97 nonrespondents with probabilities proportional to these weights was used in order to reduce the overall unequal weighting effects for the sample. These weights are the basis for computation of the B&B:93/03 analysis weights.

Table 5. Sampling rates and sample allocation for subsampling B&B:93/97 nonrespondents: 2003

Subsampling strata	B&B:93/03 advance tracing status	Control of base-year institution	B&B:93/94 response status	Number eligible	Sampling rate	Subsample size
Total				11,150		10,440
B&B:93/97 respondents				10,090	1.000	10,090
B&B:93/97 nonrespondents				1,070		360
1	Student located	Public	Nonrespondent	170	0.330	60
2		Public	Respondent	120	0.330	40
3		Private[1]	Nonrespondent	90	0.330	30
4		Private[1]	Respondent	90	0.330	30
5	Other contact located, student not located	Public	Nonrespondent	30	0.207	10
6		Public	Respondent	60	0.244	10
7		Private[1]	Nonrespondent	20	0.500	10
8		Private[1]	Respondent	30	0.260	10
9	Neither student nor other contact located	Public	Nonrespondent	180	0.330	60
10		Public	Respondent	110	0.311	40
11		Private[1]	Nonrespondent	120	0.500	60
12		Private[1]	Respondent	60	0.222	10

[1] For the purpose of subsampling B&B:93/97 nonrespondents, private, not-for-profit and private, for-profit institutions were combined due to small cell sizes.
NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

2.2 Data Collection Design

This section describes the procedures implemented in the full-scale data collection effort for B&B:93/03. Topics discussed include the methods used to locate sample members and to inform them about the study, focusing on the development of a study website and other notification materials. A unique feature of this data collection effort is that a single web-based data collection instrument was developed to be administered to respondents in three modes: self-administered via the Internet, with a trained interviewer over the telephone using CATI, and with a trained interviewer in person using computer-assisted personal interviewing (CAPI). Other topics discussed include training of data collection staff, respondent incentives, and data collection systems.

2.2.1 Pre-Data-Collection Activities

2.2.1.1 Student website

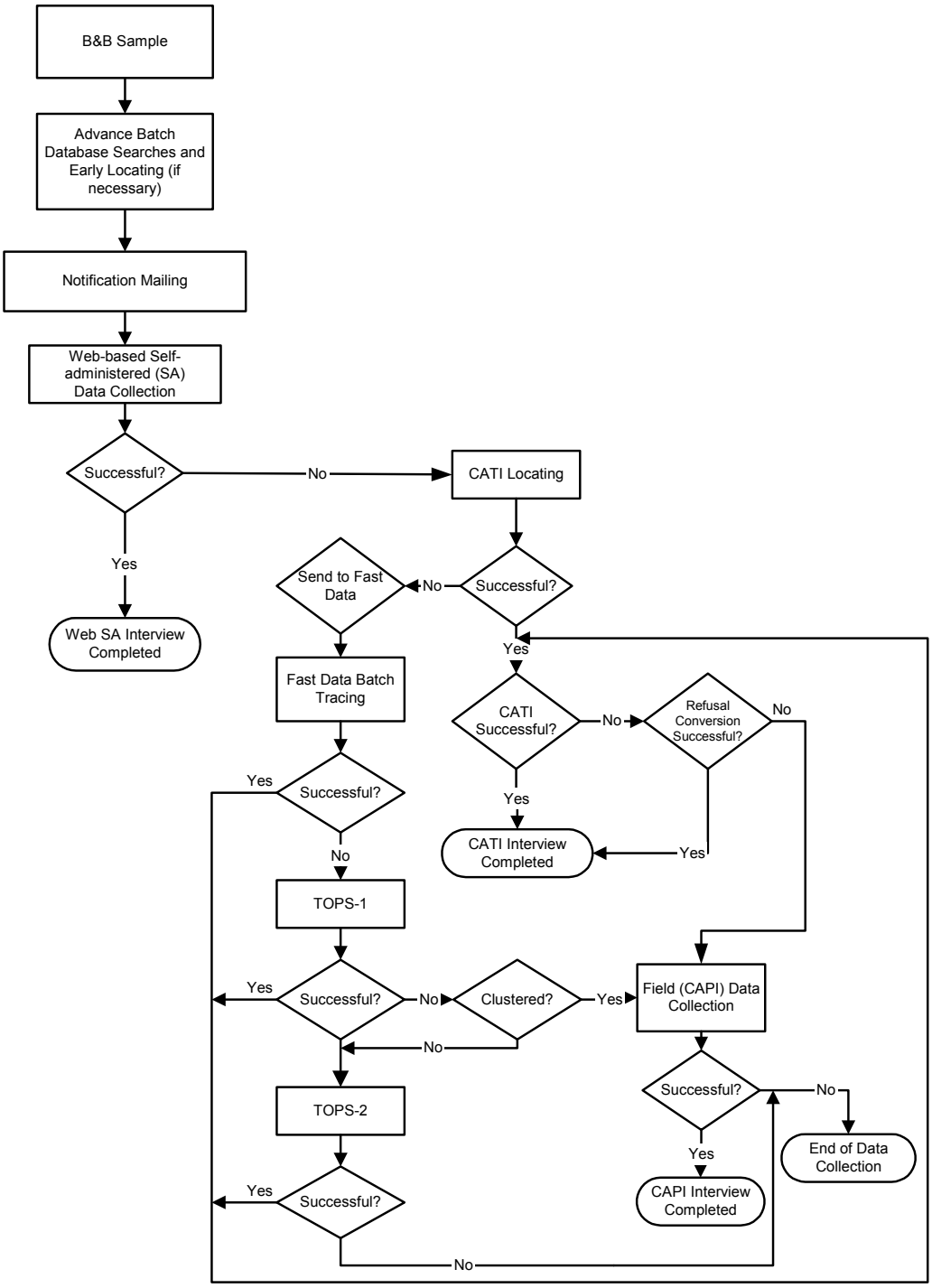
A critical element of B&B:93/03 was the design and implementation of a study website from which sample members could update address information and access the self-administered interview. The website also provided contact information for the study's Help Desk and project staff, links to the National Center for Education Statistics (NCES) and RTI websites, and information about the B&B:93/03 study, such as the history of the study and a summary of findings from prior interviews. Because the website address was included with all mailings to sample members, it could be accessed by sample members beginning with the first prenotification mailing.

The B&B:93/03 website was designed in accordance with NCES web policies. A two-tier security approach was used to protect all address and interview data collected through the website. At the first tier, sample members were required to log on to the secure areas of the website using a unique and randomly assigned Study ID sent by mail. In addition, access to the interview required a password that consisted of information from a prior interview that only the respondent would know. At the second tier of security, data entered on the B&B:93/03 website—both contact information and interview responses—were protected with Secure Socket Layer (SSL) technology, ensuring that only encrypted data were transmitted over the Internet.

2.2.1.2 Advance locating

Advance tracing activities for the full-scale B&B:93/03 were conducted prior to the start of data collection so that updated locating information could be obtained for full-scale sample members (see figure 1). In the fall of 2001, as part of the B&B:93/03 field test, RTI sent the full-scale sample for batch searches conducted on databases from the Department of Education's Central Processing System (CPS), the National Change of Address (NCOA), TransUnion's credit information, Telematch, and ComServ's Death Information System (DIS) databases. In the fall of 2002, RTI resubmitted the full-scale sample to CPS, NCOA, and Telematch to provide more current tracing information. For many sample members, these searches either confirmed existing information or yielded new locating data. Cases for which batch locating efforts were unsuccessful were sent to RTI's Tracing Operations (TOPS) unit for further tracing.

Figure 1. B&B:93/03 data collection process



NOTE: CATI = Computer-assisted telephone interview; CAPI = Computer-assisted personal interview; TOPS = Tracing Operations.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

2.2.1.3 Notification materials

The primary goal of RTI's full-scale tracing plan was to make early contact with sample members to evaluate the accuracy of contact information obtained from advance tracing efforts and provide information that would rekindle sample member interest in B&B:93/03. RTI designed several pre-data-collection mailouts to accomplish these goals. All data collection mailout materials are presented in appendix A.

Prenotification mailing. The first step in contacting B&B:93/03 sample members was to send a prenotification mailing 3 months prior to the start of data collection. The prenotification mailing contained materials describing the study, including the study leaflet that described the purpose and history of B&B:93/03. It also included a letter with a Post-it® note attached that listed the B&B:93/03 web address and toll free number, and an address update sheet with a postage-paid return envelope. All materials provided sample members with the website address and a toll-free number to contact the B&B:93/03 project director with any questions or concerns.

If the prenotification mailing was returned as "undeliverable," any new forwarding information provided by the post office was entered into a locator database. Up to two additional mailings were sent to the next "best-known" mailing address for each case in which a mailing was returned as undeliverable without a forwarding address. Cases for which a second remail was returned as undeliverable were sent to TOPS for advance tracing in order to obtain a good address for the subsequent mailings (described below).

2.2.1.4 Notification Mailings

Two weeks before the start of data collection, RTI sent a postcard to all B&B:93/03 full-scale sample members. The postcard reminded sample members that data collection was about to begin and alerted them that a package containing important information about the B&B:93/03 interview would be sent in about a week. The postcard also provided the B&B:93/03 web address and toll-free number.

One week before the start of data collection, RTI mailed lead letter packets to sample members. The mailing included a personalized letter, a Post-it® note with the study website address and toll-free number, a study leaflet, an address update sheet with a postage-paid return envelope, and a magnetic picture frame with an insert displaying the sample member's Study ID and password. The purpose of this mailing was to inform sample members that data collection was starting and to provide information on how to complete the interview.

Both the postcard and lead letter were sent to sample members in three successive waves. The sample was divided evenly into three groups to ensure that the Telephone and Internet Operations (TIO) unit was not overburdened at the start of CATI data collection.

2.2.2 Interview Design

The first step in creating the B&B:93/03 interview was to build upon the data elements used in prior surveys with the B&B:93 cohort. The data elements were developed with input

from the study's Technical Review Panel (TRP) (see appendix B for a list of members) as well as from NCES and other Department of Education staff. See appendix C for the final set of data elements used in the B&B:93/03 survey.

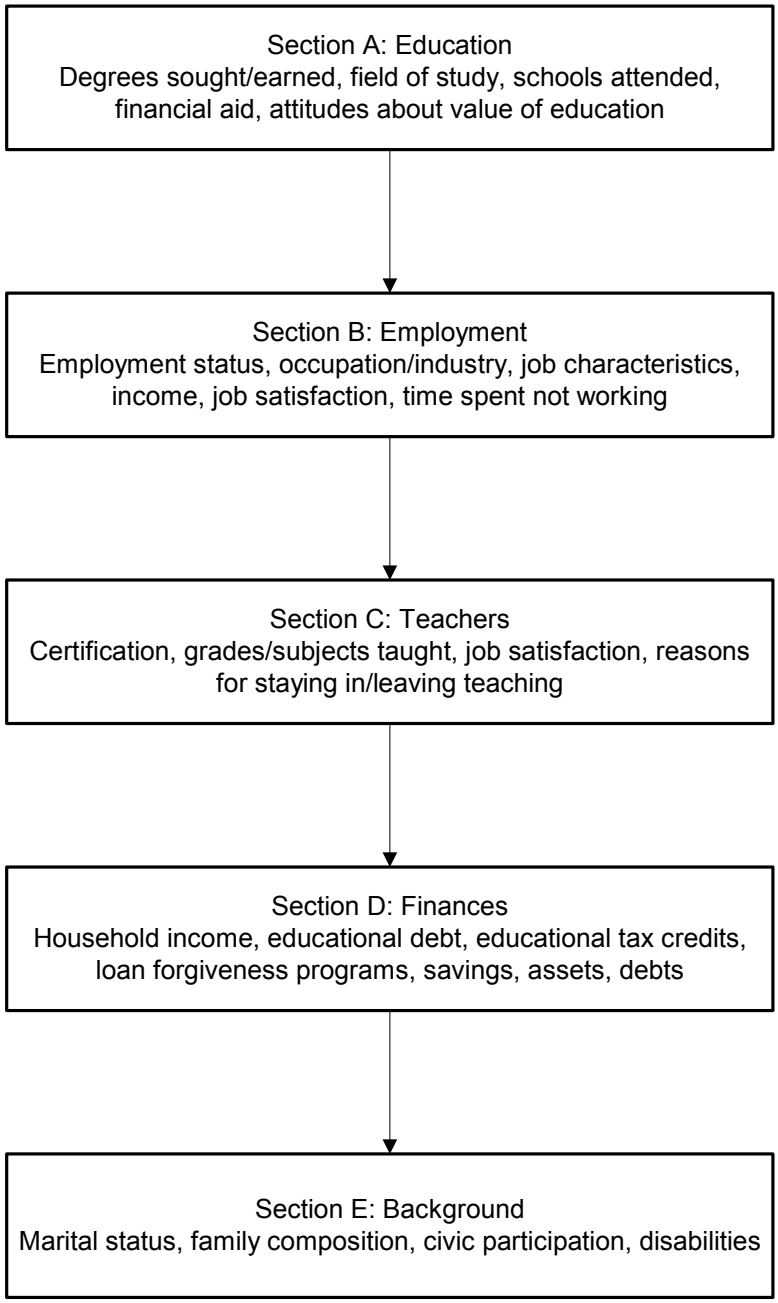
The B&B:93/03 interview was first developed for implementation in the field test and was then revised for full-scale administration based on recommendations from NCES and the TRP. Instrument specifications defined the structure of sections, variable names and definitions, skip patterns, out-of-range limits, and, when necessary, item verifications.

Figure 2 shows the progression and topics covered in the interview sections. The education section gathered information on any postsecondary education obtained since the last interview (B&B:93/97), including formal degree programs (undergraduate and graduate), credit and noncredit coursework, and courses to obtain or maintain certification and licensure. The employment section collected information on employment status, characteristics of respondents' careers, and information concerning any time spent out of the workforce. The next section for teachers and those considering teaching contained questions about teaching (kindergarten through grade 12) to monitor movement into and out of the teacher pipeline and to gather information on teacher preparation, job characteristics, and job satisfaction. To better understand the costs and benefits of obtaining a baccalaureate degree, the finance section contained questions focusing on income, assets, debts, and savings, as well as education loan burden. The final section obtained student demographic characteristics, focusing on marital status and family composition, volunteerism and political activism, as well as disability status. Facsimiles of the B&B:93/03 interview and the previous interviews used for this cohort are provided in appendix D.

A single, web-based instrument was designed and programmed for the B&B:93/03 interview for use in three modes of data collection: as a self-administered web interview, telephone interview (CATI), and field interview (CAPI). B&B:93/03 sample members could access the interview directly from the study website by entering the unique Study ID and password provided to all sample members. Telephone interviewers accessed the web interview through the case management system in RTI's TIO unit. Field interviewers accessed the interview through a case management system installed on each field laptop, and the interview was run from the laptop's own local web server and database engine with interview data downloaded nightly.

Self-administered respondents and interviewers were guided through the interview questions depending on skip logic that used answers to previous questions and preloaded information from previous interviews. When necessary, pop-up messages appeared with text, intended to clarify inconsistent or out-of-range responses or to convert item nonresponse. Coding systems were implemented to standardize the categorization of major, occupation and industry, postsecondary institutions attended by respondents, and, for respondents who taught, elementary and secondary schools in which they worked.

Figure 2. Progression of the B&B:93/03 interview



SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

Since the same instrument was used for all modes of administration, it was important to ensure that self-administered respondents and interviewer-administered respondents could respond to the same stimulus. Several steps were taken to achieve this goal. First, a link to a separate help text window (the same help text was available to interviewers and respondents) was provided on every page so that respondents and interviewers were able to view definitions of terms and clarifications of question intent. Interviewers were instructed to use the help text when needed to make sure that all respondents received the same information when they had questions about a particular item. Another difference due to the mode of administration was that web

respondents were able to read the list of response options, and CATI respondents could not. To remedy this discrepancy, interviewer instructions were displayed (for CATI/CAPI cases only) at the top of each screen indicating whether response options were to be read to respondents. A benefit to CATI/CAPI respondents that web respondents did not share was that interviewers could indicate the progress of the interview and encourage CATI/CAPI respondents to continue. Thus, a progress bar was displayed for self-administered respondents. The progress bar indicated how much of an interview section had been completed and, overall, how many sections had been completed.

2.2.3 Data Collection Activities

2.2.3.1 Staff training

Various types of data collection staff were used for the B&B:93/03 full-scale data collection, including tracing specialists, supervisors and monitors, Help Desk agents, telephone interviewers, and field interviewers. Specialized training sessions were conducted for each of these groups. A sample training agenda and table of contents from a training manual are provided in appendix E. Each training session covered an overview of the study, review of confidentiality requirements, a demonstration interview, question-by-question review of the instrument, as well as hands-on practice with the tracing module, instrument, and coding systems. In addition, each training session contained specialized instruction for each job, as described below.

- *Tracing specialists* received instruction on project-specific tracing protocols for tracing the sample members, as well as on the most effective tracing sources.
- *Supervisors and monitors* received instruction on project specific supervision and monitoring guidelines.
- *Help Desk agents* received training on answering questions about the study, as well as technical questions from sample members, and were trained to document each call made to the study hotline.
- *Telephone interviewers* received information on the content of the interview, as well as on gaining cooperation from sample members, parents, and other contacts, and techniques for refusal avoidance and addressing the concerns of reluctant participants.
- *Field interviewers* received information on the content of the interview, as well as training on field-specific operations, including the field case management system and field tracing procedures.

2.2.3.2 Early response incentive

In an effort to increase response rates and reduce the costs associated with telephone interviewing, all respondents were offered an incentive (a $20 check) for completing the self-administered web interview within the first 3 weeks of data collection. Sample members were notified of this opportunity in the lead letter mailing. Only web self-administered interviewing was possible during this 3-week period.

2.2.3.3 Help Desk

A Help Desk was available throughout data collection to assist sample members who had questions or problems accessing and completing the self-administered interview. The Help Desk was set up to receive inbound calls to a toll-free number, which had been provided to sample members in the study's notification materials. Help Desk staff were trained to assist sample members with both technical questions and questions about the study, as well as to conduct interviews. If sample members called the Help Desk because they had difficulty completing the self-administered interview, Help Desk agents encouraged sample members to complete the interview over the telephone at that time.

Help Desk agents used a specially-designed application to systematically document all calls from sample members. The Help Desk application also provided the following:

- information needed to verify a sample member's identity;
- login information allowing a sample member to access the web interview; and
- reports on the types and frequency of problems experienced by sample members and their resolution status.

2.2.3.4 Interviews

Self-administered interviews. B&B:93/03 introduced self-administered web interviewing as an option to sample members for the first time. There were many unique features and benefits of the self-administered web interview:

- respondents could complete the survey at a convenient time;
- the interview could be completed at any location with computer access (home, work, library, school, etc.);
- respondents were able to break off the interview and resume at another time;
- security features included password-protected login, encrypted data transmission, and automatic logout after idle periods;
- the CATI case management system (CMS) controlled access to active cases so that respondents in the process of completing the self-administered web interview would not be called by an interviewer; and
- if they encountered any problems with the self-administered interview, sample members were offered the option of completing the survey with an interviewer by calling the Help Desk.

Sample members were notified of this data collection option during the initial notification mailing. For the first 3 weeks of data collection, only web interviews were completed unless a respondent called the Help Desk with questions about completing the telephone interview.

Telephone interviews. CATI began at the end of the 3-week web interviewing period. The CATI software used an embedded automated call-scheduler to assign and deliver cases to interviewers. This system allowed calls to be scheduled on the basis of case priority and time of

day. Case assignments made by the scheduler maximized the likelihood of contacting and interviewing sample members by using information from prior call outcomes. The call-scheduler also prevented CATI calls from being made to cases that were currently in progress on the web or had recently been completed.

Limited tracing was conducted by telephone interviewers when sample members could not be located at a known telephone number during CATI. Telephone interviewers used Fast Data and directory assistance services, as well as locating information for the sample member provided by contacts during an earlier interview. Cases that could not be located using any of the existing address information were sent for intensive tracing in RTI's TOPS unit. Cases not located in TOPS were either sent to the field for locating and interviewing or were returned to TOPS for additional intensive tracing (TOPS2).

Field interviews. B&B:93/03 full-scale field CAPI activities began approximately 4 months following the start of CATI. Using the best known address for each case, RTI's Geographic Information System (GIS) program conducted an analysis of the B&B:93/03 sample to determine the 30 geographic areas with the highest density of B&B:93/03 sample members. Based on that analysis, field interviewers (FIs) were hired for each cluster to work any nonrespondent case residing within a 50-mile radius of the cluster center.

Field interviews were conducted by the local FI either in person or by telephone. A Field Management System (FMS) located on each Fis laptop assisted with caseload management. The FMS also updated reports that informed project staff of the progress of the field interviewing effort. Once assigned to the field, cases could not be accessed by CATI interviewers but could still be completed as a self-administered interview over the Internet.

2.2.3.5 Nonresponse incentive

Use of incentives for nonrespondents has been shown to be effective in increasing response rates while containing data collection costs (Riccobono et al. 2001; Wine et al. 2002; Wine et al. 2004). A nonresponse incentive was offered to three types of nonrespondents: those who initially refused the interview, those for whom intensive tracing yielded a good mailing address but no telephone number, and those identified as "hard to reach," that is, those with 15 or more call attempts and those with whom contact had been established but no appointment scheduled. All cases assigned to FIs were also treated as nonrespondent cases. The nonrespondent incentive mailing consisted of a letter tailored to the specific type of nonrespondent (see appendix A) and an offer to receive a $20 check upon completion of the interview.

2.3 Data Collection Systems

2.3.1 Instrument Development and Documentation System (IDADS)

The Instrument Development and Documentation System (IDADS) was a combination web and Visual Basic (VB) environment in which project staff developed, reviewed, modified, and communicated changes to specifications, code, and documentation for the B&B:93/03 instrument. All information relating to the B&B:93/03 instrument was stored in a Structured

Query Language (SQL) Server database and was made accessible through web browser and Windows VB interfaces. There were three modules within IDADS: specification, programming, and documentation.

2.3.1.1 Specification module

The IDADS specification module provided tools and graphical user interfaces for creating, searching, reviewing, commenting on, updating, importing, and exporting information associated with instrument development. A web interface provided access to the instrument specifications for project staff at MPR and NCES.

2.3.1.2 Programming module

Once specifications were finalized, the programming module within IDADS produced hypertext transfer markup language (HTML), Active Server Pages (ASP), and JavaScript template program code for each screen based on the contents of the SQL Server database. This output included question wording, response options, and code to write the responses to a database, as well as code to automatically handle such web instrument functions as backing up and moving forward, recording timer data, and linking to context-specific help text. Programming staff edited the automatically-generated code to customize screen appearance and to program response-based routing.

2.3.1.3 Documentation module

The documentation module contained the finalized version of all instrument items, their screen wording, and variable and value labels. Also included were the more technical descriptions of items such as variable types (alpha or numeric), information regarding to whom the item was administered, and frequency distributions for response categories. The documentation module was used to generate the instrument facsimiles and the deliverable Electronic Codebook (ECB) input files.

2.3.2 Integrated Management System (IMS)

All aspects of the study were controlled using an Integrated Management System (IMS). The IMS was a comprehensive set of desktop tools designed to give project staff and NCES access to a centralized, easily accessible repository for project data and documents. The B&B:93/03 IMS consisted of several components: the management module, the Receipt Control System (RCS) module, and the instrumentation module.

2.3.2.1 Management module

The management module of the IMS included tools and strategies to assist project staff and the NCES project officer in managing the study. All management information pertinent to the study was located there, accessible via the Web, and protected by SSL encryption and password-protected login. Available on the IMS were the current project schedule, monthly progress reports, daily data collection reports and status reports (generated by the RCS described below), project plans and specifications, key project information and deliverables, instrument specifications, staff contacts, the project bibliography, and a document archive. The IMS also

had a download area from which staff at MPR and National Center for Education Statistics (NCES) could retrieve large files when necessary.

2.3.2.2 Receipt Control System (RCS)

The RCS was an integrated set of systems that monitored all activities related to data collection, including tracing and locating. Through the RCS, project staff were able to perform stage-specific activities, track case statuses, identify problems early, and implement solutions effectively. The RCS's locator data were used for a number of daily tasks related to sample maintenance. Specifically, the mailout system produced mailings to sample members, the query system enabled administrators to review the locator information and status for a particular case, and the mail return system enabled project staff to update the locator database as mailings or address update sheets were returned or forwarding information was received. The RCS also interacted with the TOPS database, sending locator data between the two systems as necessary.

A subcomponent of the RCS, the Field Case Management System (FCMS), controlled field interviewing activities. The FCMS allowed field staff to conduct tracing and CAPI, communicate with RTI staff via electronic mail, transmit completed cases, and receive new cases.

2.3.2.3 Instrumentation module

The instrumentation module managed development of the multimode web data collection instrument within IDADS. Developing the instrument with IDADS ensured that all variables were linked to their item/screen wording and thoroughly documented.

2.3.3 The Variable Tracking System (VTS)

The central mechanism for constructing input files for the NCES ECB was a software application called the Variable Tracking System (VTS). The VTS tracked and stored documentation for both interview and derived variables required for the ECB and NCES' Data Analysis System (DAS). This included weighted and unweighted variable distributions, variable labels and codes, value labels, and a text field describing the development of each variable and the programming code used to construct it. Input files for the ECB and DAS systems were automatically produced by the VTS according to NCES specifications.

Chapter 3
Data Collection Outcomes

Full-scale data collection for the 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03) required several steps involved with locating sample members and their completion of either a self-administered, telephone, or in-person interview. This chapter presents the outcomes of the data collection (including interview rates, both overall and by mode) and burden on respondents. It also assesses the effectiveness of the data collection strategies used in locating, contacting, and interviewing sample members.

3.1 Contacting and Interviewing Outcomes

Contacting and interviewing results for the B&B:93/03 full-scale data collection are presented in figure 3. Prior to the start of data collection, 10,440 sample members were considered eligible to participate in the B&B:93/03 interview, including 360 nonrespondents to the B&B:93/97 interview. Of the initial sample, 0.4 percent was found to be either study ineligible or deceased. Ninety three percent of sample members were located (i.e., RTI interviewers reached the sample member in his or her household), and the remaining 7 percent could not be located throughout data collection.

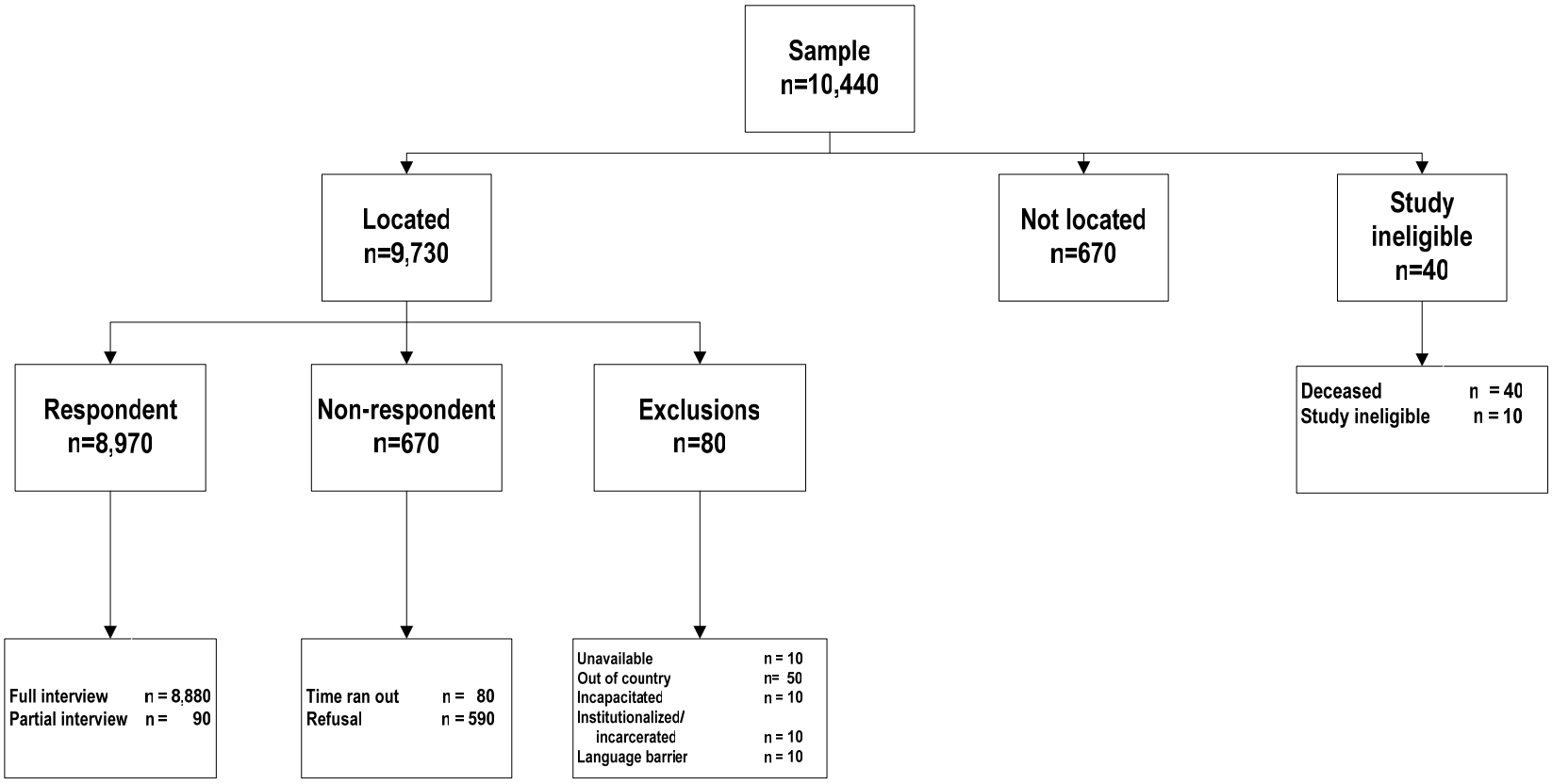
Less than 1 percent, of those located sample members were excluded from data collection because they were found to be out of the country, institutionalized, incarcerated, physically or mentally incapacitated, or otherwise unavailable for the duration of the data collection period. Another 6 percent refused to participate either directly or through a gatekeeper. Among the B&B:93/03 sample members who were eligible for participation, 8,970 were interviewed, for an overall unweighted response rate of 86.3 percent (83.4 percent weighted response rate).

3.1.1 Interviewing Outcomes by Mode

B&B:93/03 used a web-based, multimode data collection strategy which combined self-administered, telephone, and in-person interviewing options. For the first 3 weeks of data collection, only the self-administered web option was available. Help Desk staff were trained to provide assistance with computer problems as needed by web respondents and to complete a telephone interview if the computer problems could not be resolved. At the end of the 3-week, "web-only" period, computer-assisted telephone interviewing (CATI) began with all incomplete cases. Three months following the start of CATI, selected nonrespondent cases were sent to the field for computer-assisted personal interviewing (CAPI).

Completion mode for student interviews is presented in table 6. Although the web, self-administered option was new to the B&B:93 cohort, 38 percent of respondents chose this method, whereas 57 percent of those responding completed their interview by telephone. For the first 3 weeks of data collection, sample members could only complete the B&B:93/03 interview over the Web. Any sample member who completed the interview during the 3-week web period was paid an incentive (see section 3.3).

Figure 3. Locating and interviewing outcomes



NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

Table 6. Student interview mode of administration: 2003

	Number	Percent
Total	8,970	100.0
Web	3,420	38.2
Computer-assisted telephone interview	5,070	56.5
Computer-assisted personal interview	480	5.3

NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

CATI locating and interviewing began 3 weeks following the start of web data collection and continued for almost 7 months. By the end of data collection, 49 percent of the sample had completed the B&B:93/03 interview by telephone. Although telephone interviewing did not begin until after the web early incentive period expired, about 110 telephone interviews were conducted by Help Desk staff during the incentive period if it was determined that the respondent had insurmountable computer problems, preventing completion of the web interview.

About 3 months following the start of CATI data collection, field interviewers began tracing and interviewing all interview nonrespondents whose last known address was in one of 30 geographic clusters. A total of 700 cases were sent to the field during the 3.5 months of CAPI data collection. Of those, 68 percent completed an interview, which represents 5 percent of all respondents.

Typically, in a longitudinal study, participation in a preceding interview is a good predictor of the likelihood that a sample member will participate in a future interview. For B&B:93/03, a comparison of contact and interview rates in table 7 shows that those who participated in B&B:93/97 were more likely to be both located and interviewed during the B&B:93/03 interview. More B&B:93/97 respondents were located (94 percent) and interviewed (88 percent) than nonrespondents (76 percent contacted [$z = 11.9$; $p < 0.01$] and 52 percent interviewed [$z = 14.1$; $p < 0.01$]).

Table 7. Contact and interview rates, by prior response status: 2003

	Outcome in B&B:93/03			
	Overall		Percent	Percent
Response status in B&B:93/97	Number	Percent	located	interviewed
Total	10,400	100.0	93.5	86.3
Respondent	10,050	96.6	94.1	87.5
Nonrespondent	350	3.4	76.1	51.7

NOTE: Total sample does not include those sample members who were found to be ineligible or deceased. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

3.2 Locating and Interviewing Outcomes

Effectively tracing and locating sample members was critical to the success of the B&B:93/03 data collection effort. Since the last contact with sample members occurred in 1997 or in 1994 for B&B:93/97 nonrespondents, tracing and locating were expected to be particularly difficult. Locating activities required tracing prior to data collection, additional tracing by interviewers during data collection, intensive tracing by RTI's Tracing Operations (TOPS) unit, and tracing by field interviewers. The results of each of these tracing activities are presented below.

3.2.1 Pre-Data-Collection Tracing

Tracing of the full-scale sample began in the fall of 2001 by updating any contact information collected during the B&B:93/94 and B&B:93/97 interviews. Several tracing sources were used, including the Central Processing System (CPS), which contains federal financial aid application information; the National Change of Address (NCOA) from the U.S. Postal Service; databases from Telematch; TransUnion's credit information; and the Death Information System (DIS). Before B&B:93/97 nonrespondents were stratified and subsampled, the entire full-scale sample was sent for batch tracing. Table 8 shows the record match rate for each method of batch tracing employed.

Table 8. Batch processing record match rates, by tracing source: 2003

Method of tracing	Number of records sent	Percent matched[1]
Central Processing System (2002–03)	11,070	1.7
National Change of Address	11,180	37.4
Telematch	11,180	73.5
TransUnion	11,190	99.0
Death Information System	11,080	0.4

[1]Percent match rate is based on the number of records sent for batch tracing as part of the sample stratification process. Since records were sent to multiple tracing sources, multiple records matches were possible. Because different information was required for each method of tracing, the number of records that could be sent varied by tracing method.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

Pre-data-collection tracing continued with a mailout to the 10,400 eligible sample members. By mailing prenotification materials to the best known address for a sample member, more recent information could be obtained from forwarding orders provided by the U.S. Postal Service. In addition, as part of the mailing, sample members were asked to provide an address update either on the study website or on an address update sheet returned to project staff in a self-addressed, postage paid envelope. Overall, 18 percent of sample members provided updates (table 9).

Table 9 shows the contact and interview rates for those who provided an address update by the mode used to provide the update. Self-reported address updates virtually assured contacting and interviewing the sample member. Almost all of those who updated their contact information were subsequently located (99 percent) and interviewed (95 percent). Although the study website could receive address information electronically, almost twice as many updates were received via conventional mail.

Table 9. B&B:93/03 contact and interview rates, by type of address update reply

Mode of address update	Total	Percent located[1]	Percent interviewed[1]
Total	1,910	99.4	95.0
Hardcopy	1,210	99.4	99.3
Website	690	99.4	99.3

[1] Percentages are based on the total within the row under consideration.
NOTE: Sample members who are deceased, unavailable, or incapitated are excluded. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

3.2.2 Tracing During Data Collection

During data collection, if all leads for a sample member were exhausted before the individual was located, interviewers could opt to send the case for Fast Data batch processing. Fast Data conducts a series of database searches on existing address information available for a case. A total of 2,960 cases were sent to Fast Data, 67 percent of which were returned with new information or a verification of existing information. Of those cases returned, 84 percent were located and 76 percent were interviewed.

3.2.3 Intensive Tracing

If a case could not be located through Fast Data, it was sent for intensive tracing conducted by RTI's TOPS unit. Because their tracing information was assumed to be too outdated to be useful, intensive tracing on all B&B:93/97 field test nonrespondents selected to participate in the B&B:93/03 interview occurred prior to the start of data collection. A number of locating sources were used during intensive tracing, including consumer databases, directory assistance, and internet sources.

Table 10 provides the results of the TOPS intensive tracing efforts; locating and interviewing rates for sample members not requiring TOPS tracing are provided for comparison. Each phase of TOPS tracing allowed for more in-depth tracing efforts. As shown in table 10, of the 2,610 cases traced using TOPS1 procedures, 84 percent were located, with 75 percent interviewed. The most comprehensive stage of locating activities, TOPS2, was used only when all previous tracing efforts failed and the sample member was not sent for field interviewing (only cases located in one of the 30 geographic clusters identified for CAPI were sent to the field). Of the TOPS2 cases, 57 percent could be located, with 51 percent interviewed. Clearly, compared with TOPS1 cases, TOPS2 cases were more difficult to locate ($z = 9.1$; $p < 0.01$) and, once located, more difficult to interview ($z = 6.6$; $p < 0.01$).

Sample members who were not B&B:93/97 respondents tended to be difficult to locate even after TOPS1 and TOPS2 tracing activities were completed. Of the B&B:93/97 nonrespondents sent for TOPS1, only 52 percent were located and 37 percent interviewed. Of the cases sent for TOPS2, only 40 percent were located and 36 percent interviewed.

Table 10. Contact and interview rates, by intensive tracing status: 2003

Intensive tracing status	Total	Percent located	Percent interviewed
Total	10,400	93.5	86.3
Cases requiring first level intensive tracing (TOPS1)	2,610	84.2	75.3
Cases requiring second level intensive tracing (TOPS2)	310	57.0	51.1
Cases not requiring intensive tracing	7,480	98.3	91.5

SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

3.3 Early Response Incentive

As discussed above, only the web-based, self-administered interview was available at the start of data collection. Sample members were given 3 weeks to complete the interview themselves, and those who did complete during the first 3 weeks of data collection were paid a $20 incentive. Of the 3,420 respondents who completed a self-administered interview, 47 percent completed it during this 3-week period. An additional 2 percent of respondents began the interview on the Web, but completed it in CATI or with a field interviewer (and, therefore, are counted as either a CATI or CAPI complete rather than as a web complete) during the 3-week period.

3.4 Refusal Conversion Efforts

Refusal conversion procedures were used to gain cooperation from individuals who refused to participate in the interview. When a refusal was first encountered, either because the sample member refused or because a "gatekeeper" refused on behalf of the sample member, the case was referred to a refusal conversion specialist. Refusal conversion specialists were selected from among those interviewers most skilled at obtaining cooperation and were given training in refusal conversion techniques tailored to the B&B:93/03 interview. The training emphasized how to gain cooperation, overcome objections, address the concerns of gatekeepers, and encourage participation.

Of the 10,400 sample members, 10 percent were referred to refusal conversion specialists. Among those, 49 percent were successfully converted, i.e., the interview was completed. Most of the converted interviews (60 percent) were completed by telephone, with 39 percent completed by Web and only 1 percent completed by field interviewers.

Table 11 provides a breakdown of refusal conversion rates by prior response status. Not surprisingly, B&B:93/97 respondents were less likely to refuse to participate in the B&B:93/03 interview than were nonrespondents ($z = -5.2$; $p < 0.01$). In addition, conversion rates were higher among B&B:93/97 respondents than among nonrespondents. Of the B&B:93/97 respondents who refused to participate in the B&B:93/03 interview, 51 percent were eventually completed. Of the B&B:93/97 nonrespondents who refused, only 32 percent completed the interview. Participation in prior B&B:93 cohort interviews tended to increase the likelihood of agreeing to be interviewed for B&B:93/03 ($z = 2.01$; $p < 0.05$).

Table 11. B&B:93/03 refusal and refusal conversion rates, by prior response status

Prior response status	Total	Percent ever refused B&B:93/03 interview	Percent interviewed, after refusal
Total	10,400	9.5	49.4
1997 Respondents	10,050	8.9	51.2
1997 Nonrespondents	350	26.4	32.3

NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

3.5 Nonresponse Incentive

In addition to the early response incentive described above, an incentive was used later in the B&B:93/03 data collection process to reduce nonresponse among four groups: those who initially refused to participate in the study, those who were difficult to reach by telephone, those who could not be located but for whom a contact person could be reached, and those who started the web interview but did not complete it. Sample members in the four groups were sent a personalized letter with instructions for completing the interview either by Web or by calling the study's toll-free telephone number. The letter also indicated that respondents would receive a $20 personalized check for completing the B&B:93/03 interview.

Table 12 presents the contact and interview rates for the nonrespondent groups. About half of the B&B:93/03 sample (50 percent) was eligible for a nonresponse incentive at some point during data collection, once the early response incentive period ended. Of those, 85 percent were eventually located and 55 percent interviewed. Almost all (93 percent) refusal cases were located (the remaining 7 percent of cases became refusals because of gatekeeper refusals), with a 55 percent conversion rate once the incentive was offered. Those respondents whose cases were designated as "unable to locate" were difficult to reach to offer the incentive; 73 percent of those originally classified as unlocatable were eventually located by an interviewer; and, once located, 9 percent were interviewed.

Table 12. B&B:93/03 contact and interview rates when an incentive was offered, by interview status

Interview status	Total	Percent located[1]	Percent interviewed[1]
Total	4,330	85.3	55.2
Refusal	1,320	92.7	54.5
Hard to reach	1,490	85.3	65.9
Unable to locate	300	73.2	9.4
Partial web interview	1,220	80.2	54.0

[1] Percent based on total within row under consideration.
NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

3.6 Interview Burden

The time burden associated with completion of the B&B:93/03 interview was calculated separately for each mode of data collection: self-administered, CATI, and CAPI. For the purposes of this analysis, however, CATI and CAPI timing data have been combined.

Figure 4 provides a visual representation of how the on-screen and transit times were determined. Two time stamp variables were associated with each interview question. The first, the start timer, was set to the clock time on the respondent's or interviewer's computer at the time that a particular web page was displayed on the screen. The second time stamp variable, the end timer, was set to the clock time on the respondent's or interviewer's computer at the moment the respondent or interviewer clicked the "Continue" button to submit the answers from that page.

From the two time stamp variables, an on-screen time and transit time were calculated. The on-screen time was calculated by subtracting the start time from the end time for each web page that the respondent received. The transit time was calculated by subtracting the end time of the preceding page from the start time of the current page; it includes the time required for the previous page's data to be transmitted to the server, for the server to store the data and assemble and serve the current page, and for the current page to be transmitted to and loaded on the respondent's or interviewer's computer.

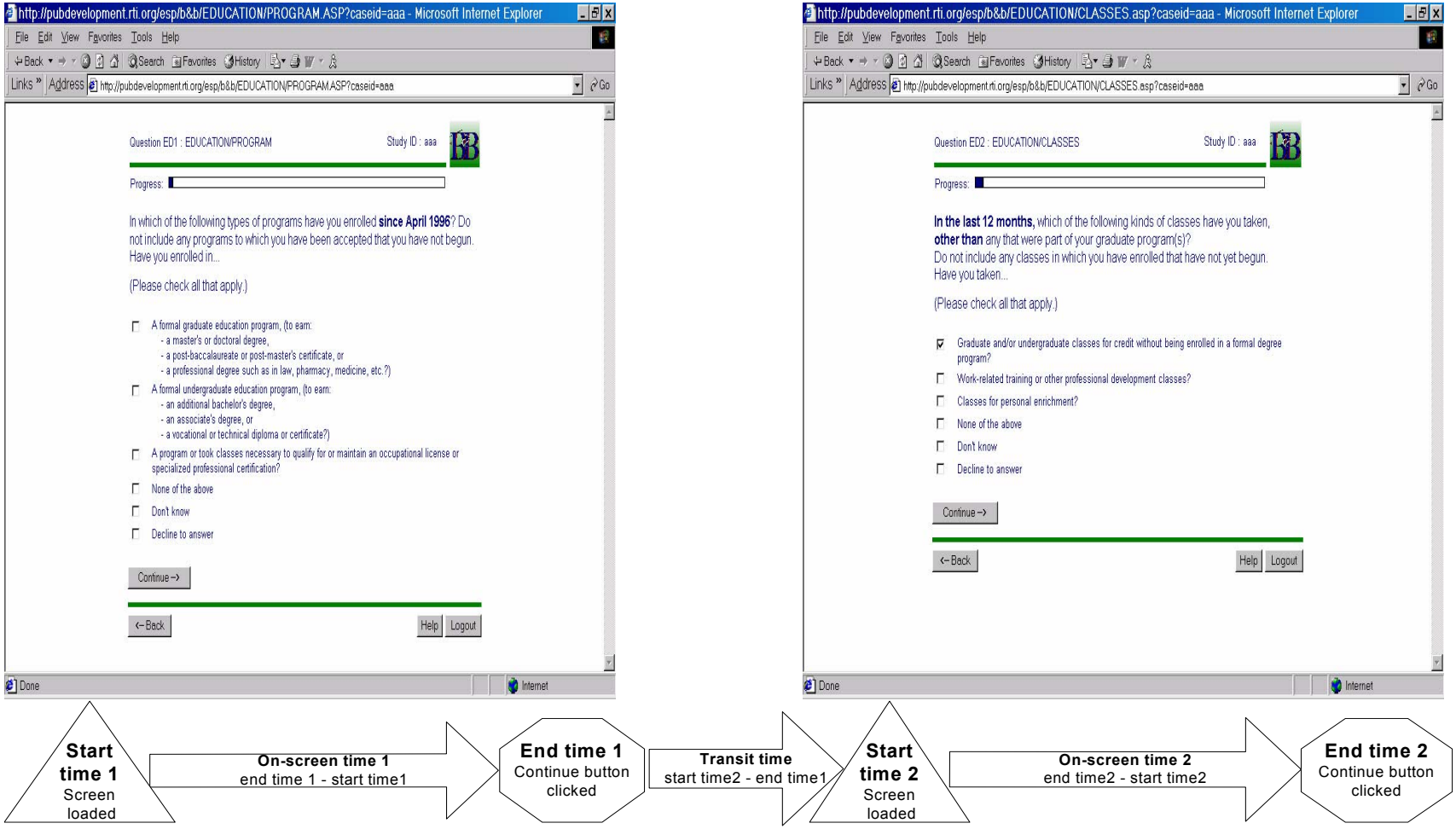
A *total on-screen time* was then calculated for all respondents by summing the on-screen times for each web page that the respondent received. For each respondent, a *total transit time* was calculated by summing all the transit times. The total on-screen and total transit times were then summed to determine the *total instrument time*.

Table 13 presents the timing results (in minutes; combining on-screen and transition times) for the entire interview and by interview section, for all respondents who completed the entire interview.[7] It also presents timing results when the interview is self-administered and when the interview is interviewer-administered (combining CATI and CAPI results). Sections are listed in the table in the order in which they were presented during the interview. Overall average time to complete the interview was just under 35 minutes, summing both on-screen and transit time.

The longest section in the interview was the employment section, which took an average of 11.6 minutes to complete. Employment questions focused primarily on two jobs—the job held in February and the job held at the time of the interview, if different. For those who earned a graduate degree since the last interview, a third set of questions was asked about the first job after degree completion (again, if different from the current job). For those unemployed at least once since the last interview, questions were asked about the duration, reasons, and specific dates for each spell of unemployment. Consequently, time in the employment section was higher for those with multiple jobs or multiple bouts of unemployment, or both.

[7] Partial interviews were excluded from the timing analysis.

Figure 4. Visual representation of on-screen and transit times



Total On-screen time = On-screen time 1 + On-screen time 2 +…+ On-screen time N
Total Transit time = Transit time 1 + Transit time 2 +…+ Transit time N
Total Instrument time = Total On-screen time + Total transit time

SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

The education section averaged 8.0 minutes to complete. Because the education section collected all education experiences, many respondents were required to answer sets of questions about enrollment in formal education programs as well as in personal enrichment and employment training courses. Those who enrolled in multiple graduate, undergraduate, and/or certificate programs looped through item sets for each program attended.

The third section of the interview focused on questions for those who entered the teacher pipeline upon completion of the bachelor's degree in the 1992–93 school year, and any new entrants to the pipeline since 1992. Those respondents who had not taught and who had no interest in or plans for teaching were skipped out of the teacher section entirely, after they answered the initial gate questions. Overall, average time in the teacher section was 3.1 minutes. However, the average time for this section for respondents in the teacher pipeline was 10.8 minutes and was only 1.0 minute for non-teachers. The background section, which collected information on family status, citizenship, political activities, volunteerism, and disability status averaged only 3.9 minutes to administer. The last section, on finance, was modified after the B&B:93/03 field test to include sets of yes/no questions rather than questions asking for specific dollar values for assets and debts. As a result, average time in the finance section decreased from 10.3 to 5.3 minutes.

Table 13 also compares average times to complete the total interview and the individual interview sections when the interview is self-administered or interviewer-administered (CATI/CAPI). There was no difference detected in the total completion time by mode of administration. Self-administered interviews averaged 34.4 minutes, and telephone interviews averaged 34.8 minutes to complete ($t = -1.48$; $p < 0.1387$).

Table 13. Average minutes to complete B&B:93/03 student interview, by interview section and mode of administration

Instrument section	All respondents	Self-administered interviews	Interviewer-administered interviews
Total interview[1]	34.6	34.4	34.8
Section A – Education	8.0	7.6	8.3
Section B – Employment	11.6	11.4	11.7
Section C – Teacher	3.1	2.7	3.4
Section C – Teacher (teachers)	10.8	10.3	11.1
Section C – Teacher (non-teachers)	1.0	1.0	1.0
Section D – Background	3.9	3.8	4.1
Section E – Finance	5.3	4.9	5.5

[1] Total interview time combines on-screen and transit times across all sections and respondents.
NOTE: Outliers were excluded from this analysis. Outliers were identified separately for each section and for the total interview, therefore, individual section times do not sum to the total interview times. An outlier was defined as any case whose completion time exceeded two standard deviations above or below the average time for a given section. Interview times are presented only for completed interviews (partial interviews were excluded.)
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

Although there was no difference by mode in the total interview completion time, there were differences within the sections. Outliers were defined separately for the total interview and for each individual section because of the nature of the self-administered interview. First, respondents were able to break off and resume the interview as was convenient for them. Second, the web-based interview automatically logged off after a predefined period of inactivity for security purposes. For these reasons, it would be possible for a self-administered respondent to have been identified as an outlier in one or more of the sections but not for the overall interview, or vice versa. Because the outliers were potentially different within the sections and for the total time, it is possible to see significant differences in section times but not for the total interview time.

More specifically, interviewer-administered interviews took significantly longer than self-administered interviews for all sections. For the education section, interviewer-administered respondents took 8.3 minutes compared with 7.6 minutes for self-administered respondents (t = -7.19; $p < 0.001$). The difference in total times for the employment section was small but significant (11.7 minutes for interviewer-administered and 11.4 minutes for self-administered; t = -3.90; $p < 0.001$). Observed differences by mode were larger for teachers. Respondents in the teacher pipeline who completed the interviewer-administered survey took 11.1 minutes, and self-administered teachers took 10.3 minutes (t = -2.99; $p < 0.05$). For non-teachers, however, the difference by mode was very small (1.03 minutes for interviewer-administered and 1.00 minutes for self-administered; t = -2.28; $p < 0.05$). In the background section, interviewer-administered respondents took 4.1 minutes compared with 3.8 minutes for interviewer-administered respondents (t = -9.76; $p < 0.001$). The finances section took 5.5 minutes for interviewer-administered respondents and 4.9 minutes for self-administered respondents (t = -13.39; $p < 0.001$).

It is likely that interviewer-administered respondents took slightly longer to complete the interview sections because respondents and interviewers were engaged in a conversation, and respondents had to wait for interviewers to read the entire question and response options (depending on the nature of the screen and the interviewer instructions[8]). Self-administered respondents, however, could read and respond to interview questions more quickly because they were able to read the entire screen at once.

Table 14 shows the total interview time broken into its components: time on-screen and time in transit. CAPI respondents have been excluded from this analysis since the CAPI was housed locally on each interviewer's laptop, which made transit times virtually instantaneous. Overall, self-administered respondents had a greater average total transit time (10.0 minutes) than did CATI respondents (5.7 minutes; t = 34.7; $p < 0.0001$).

However, when transit time is removed from the total interview time, average on-screen time for self-administered respondents (24.4 minutes) is actually significantly *less* than for CATI respondents (29.1 minutes; t = –23.05; $p < 0.0001$). At the end of the survey, a short debriefing section asked questions about users' experiences in completing the web survey. As part of the

[8] To minimize mode differences and ensure that all respondents were exposed to the same information, interviewer instructions were included on every form of the questionnaire for CATI and CAPI interviews. These instructions indicated to interviewers how to handle response options (e.g., whether the response options should be read aloud or not).

debriefing section, self-administered respondents were asked which type of internet connection they used to access the survey. Table 15 presents the average total interview times and transit times by type of internet connection, as well as the percent of total interview time accounted for by transit time.

About 20 percent of self-administered respondents completed with a dial-up connection via modem, and about 35 percent of self-administered respondents completed the survey with a fast connection (including cable modem, DSL, ISDN, LAN, etc.) Dial-up modem users took longer to complete the total interview (41.0 minutes; $t = 24.6$; $p < 0.0001$) and had longer transit times (16.1 minutes; $t = 46.6$; $p < 0.0001$) than did users with a fast connection (30.0 minutes to complete the interview and 6.3 minutes for transit time). Likewise, the percent of the total interview time taken by transit was longer for dial-up users than for users with a fast connection (39 percent compared with 21 percent, respectively; $z = 10.1$; $p < 0.01$).

Table 14. Average on-screen and transit times in minutes, by response mode: 2003

Instrument section	Average total time	Average total on-screen time	Average total transit time
All web and CATI[1] respondents	34.6	27.4	7.2
Web respondents	34.4	24.4	10.0
CATI respondents	34.8	29.1	5.7

[1]CATI = Computer-assisted telephone interview. Computer-assisted personal interviewing (CAPI) cases were excluded from this analysis.
NOTE: Times are presented separately for time on-screen and time in transit. Interview times are presented only for completed interviews (partial interviews were excluded). Outliers were identified separately for each section, and for the total interview. An outlier was defined as any case whose completion time exceeded two standard deviations above or below the average time for a given section. Outliers were also excluded from this analysis.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

Table 15. Average minutes to complete B&B:93/03 student interview, by interview and transit time, and internet connection speed

Internet connection speed	Average total time	Transit time	Percent total time in transit	Percent of cases
Dial-up modem	41.0	16.1	39.2	18.6
Fast connection	30.0	6.3	21.0	36.1
Cable modem	29.1	5.7	19.5	16.5
Digital Subscriber Line (DSL)	29.6	6.0	20.2	9.7
ISDN	30.8	8.2	26.6	0.7
Corporate LAN (TI,T3, etc.)	32.0	7.7	24.1	9.2
Don't know	38.0	12.1	31.8	8.4
Other	39.8	13.0	32.7	0.7

NOTE: At the end of the interview, a debriefing section was included that asked questions about self-administered respondents experiences in completing the web survey. Data presented here are based on the self-administered respondents who answered the debriefing questions. Fast connection is the average interview time of respondents with a Cable Modem, Digital Subscriber Line, ISDN, or Corporate LAN. Average total time is sum of on-screen and transit times.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

3.7 Staff Burden and Effort

3.7.1 Help Desk

To better understand the issues encountered by sample members attempting the web interview, a software program was created to document each Help Desk incident that occurred during the field test. For each incident, Help Desk staff confirmed contact information for the respondent and recorded the respondent's Study ID, a description of the problem and resolution, its status (pending or resolved), and the approximate time it took to assist the caller.

Help Desk staff were trained both to work the Help Desk hotline and to conduct telephone interviews when needed. Help Desk operators also responded to e-mail messages sent to the project e-mail account and to voice mail messages left by sample members when the Call Center was closed. Each of these types of contacts was entered into the Help Desk system and documented. Almost 10 percent of sample members called the Help Desk during data collection, most of whom (89 percent) only needed to contact the Help Desk one time.

Table 16 provides detail on the types of incidents encountered for cases that required Help Desk assistance. The majority of incidents (56.4 percent) recorded by the Help Desk were from sample members requesting their Study ID or password, or both, with 9 percent of the calls asking about browser settings and computer problems. Program errors, reports of perceived logic problems, and reports of website unavailability together accounted for only 4 percent of Help Desk calls. Almost 20 percent of sample members called the Help Desk to complete the interview over the telephone. (Those attempting to complete the self-administered interview using a dial-up modem were encouraged to complete a telephone interview to minimize the time required to participate.) The remaining 9 percent of Help Desk calls were for other problems not otherwise classified.

Table 16. Response pattern, by Help Desk incident type: 2003

Type of incident	Total incidents recorded[1]	Percent of total incidents
Total cases with an incident	1,000	100.0
Study ID/password	560	56.4
Browser settings/computer problems	90	8.5
Program error	20	2.1
Routing/skip problems	#	0.2
Website unavailable	10	0.7
Question about study/instrument	30	3.3
Complete interview by telephone	200	19.7
Other problems, not classifiable	90	9.0

Rounds to zero.
[1] Multiple incidents were possible for each sample member.
NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

3.7.2 Interviewer Hours and Number of Calls

Telephone interviewing required over 15,440 telephone interviewer hours, exclusive of training, supervision, monitoring, administration, and quality circle meetings. The average time spent per completed interview was 3.09 hours. Since the average time to administer the interview was 35.3 minutes for CATI cases, the large majority of interviewer time was spent in other activities. The bulk of time was devoted to locating and contacting the sample member. Although a small percentage of noninterview time was required to bring up a case, review its history, and close the case (with the appropriate reschedule, comment, and disposition entry when completed), the bulk of time was devoted to locating and contacting the sample member.

Table 17 shows the number of telephone calls made per case, overall, and by prior response status. About 252,350 telephone calls were made during data collection, excluding those made by field interviewers, with an average of 24.3 calls made per sample member. Those interviewed were called 20.6 times, on average, less than half the average number of calls made to those not interviewed (47.3) ($t = 18.9$; $p < 0.0001$). Sample members who were nonrespondents in 1997 were called an average of 32.3 times, compared with the average of 24 calls to those who were respondents in 1997 ($t = 3.7$; $p < 0.001$).

Table 17. Average calls per case, by interview status and prior response status: 2003

	Number of cases	Number of calls	Average calls per case
Total	10,400	252,350	24.3
Interview status B&B:93/03			
Respondent	8,970	184,780	20.6
Nonrespondent	1,430	67,580	47.3
Interview status B&B:93/97			
Respondent	10,050	240,980	24.0
Nonrespondent	350	11,380	32.3

NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

Interview nonresponse is an increasing problem for CATI and CAPI studies, affecting the cost of data collection and the quality of the resulting data. Call screening devices, such as telephone answering machines, Caller ID, call-blocking, and privacy managers, help sample members avoid unwanted telephone calls, but they can also affect the representativeness of data, lower study response rates, and increase project costs by requiring additional call attempts and interviewer time.

Of the 10,400 sample members, 70 percent had at least one answering machine event. An average of only 4.1 calls was required to obtain an interview in cases in which no answering machine was reached during the course of contacting the respondent, compared with 32.9 calls in cases in which an answering machine was reached at least once. The 30 percent of cases not reaching an answering machine required significantly fewer calls than those reaching an answering machine at least once ($t = -58.3$; $p < 0.0001$). In addition, cases with no answering

machine events had a much lower rate of ever refusing (2.9 percent) than did cases with one or more answering machine events (12.2 percent, $\chi^2 = 222.2$; $p < 0.0001$).

3.8 Conclusion

The B&B:93/03 data collection offered sample members a web, self-administered interview option for the first time, with about one-third of the sample completed the interview by Web. Despite the 6 years since the last contact with the cohort, tracing and locating efforts for B&B:93/03 were successful, with most of the sample located. Interviewing was successful once a sample member was located—over 92 percent of those located were interviewed. Use of incentives improved response rates when used early in data collection with web respondents and later in data collection with refusals and hard-to-reach cases.

Comparison of interview times by mode showed that, while interviewer-administered interviews and self-administered interviews took about the same amount of time overall, there were significant differences in how that time was spent. Compared to self-administered interviews, interviewer-administered interviews required more on-screen time to read questions and record responses. In contrast, self-administered interviews required more time in transit than did interviewer-administered interviews, primarily due to slower modem connections.

Throughout data collection, interviewers spent the majority of their time tracing and locating sample members. The prevalent use of answering machines made reaching B&B:93/03 sample members difficult.

This page left intentionally blank.

Chapter 4
Evaluation of Data Quality

The full-scale 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03) used a web-based instrument that could be either self-administered by sample members or administered to sample members by an interviewer. As part of data collection, a number of evaluations to assess the quality of the data collected by the B&B:93/03 instrument were designed. These evaluations were conducted in three major areas, each of which can impact data quality: the usability of the instrument, the effectiveness of the instrument as a multimode interview, and the effectiveness of the data collection design. The results of each evaluation are presented separately below.

4.1 Usability of the Instrument

Developing a functional web survey for B&B:93/03 meant developing a usable application. "Usability" refers to the ease with which users can work with an application to easily and quickly attain their objectives. In the context of the B&B:93/03 field test interview, users were defined as the sample members, and their objective was to complete the survey without undue burden. To the extent that the web survey is not "user friendly," data quality could be adversely affected, resulting in lower response rates and higher break-off rates. During the B&B:93/03 field test, several steps were taken to ensure usability of the field test instrument, including usability testing, evaluating two types of on-screen motivators, and development of effective on-screen help text and coding systems. Evaluating usability across modes, when appropriate, further ensured that usability was maintained for both types of users, that is, for both sample members and interviewers. Based on findings from the field test, design changes were implemented prior to the start of full-scale data collection (Wine et al. 2004).

4.1.1 Help Text

Help text was available for every web screen of the B&B:93/03 instrument. Help text screens displayed instructions on how to enter responses, the type of information requested, and definitions of words or phrases within an item. In addition, there were general help screens available that provided information on the type of internet browser to use and how to answer the survey questions (i.e., how to use a check box, drop-down box, or radio button). On every help text screen, a toll-free number to the B&B:93/03 Help Desk was provided. Counters were used to determine the number of times that each help screen was accessed. Items with high rates (in excess of 10 percent) of help text access indicated that web respondents or computer-assisted telephone interviewing (CATI)/computer-assisted personal interviewing (CAPI) interviewers needed additional information about the question before giving a valid response.

For most screens in the B&B:93/03 interview, help text usage rates were consistently under 1 percent. Only one item, GRAID1 (*Which of the following types of financial aid have you received to help cover educational expenses for your [fill degree type]?*), showed a help text access rate in excess of 10 percent. For this item, a lengthy list of options, from student loans to personal loans and tuition waivers, was provided, and respondents were to select all options that

applied. The item was administered to only the 260 respondents to whom it applied, and 55 percent of the total help text hits for this screen were made by web respondents.

4.1.2 Coding Systems

The B&B:93/03 web instrument included tools that allowed online coding of literal responses for occupation, industry, major/field of study, and area of licensure or certification. When self-administered by B&B sample members, help text and limited supporting text on screen are available to assist with online coding. When administered as a telephone or in-person interview, interviewers can take advantage of the availability of the respondent to clarify coding choices at the time the coding was performed, thereby improving data quality. Interviewers use probing techniques to facilitate the online coding process.

Throughout data collection, coding experts examined samples of each set of coding results for completeness and for the correctness of codes selected by respondents (self-administered interviews) and interviewers (CATI and CAPI interviews). These expert coders determined whether the selected code was the appropriate code, whether a different code should have been assigned, or whether a string was too vague to be evaluated for recode. For the full-scale data collection, up to 50 percent of all codes were selected for evaluation. Verbatim strings and codes were provided to coders as a single data file, irrespective of the mode of data collection (self or interviewer-administered interview).

Table 18 shows the results of the recode analysis for each coding system. Overall, almost 73 percent of the original codes were appropriate given the verbatim string provided by respondents. About 4 percent of the strings were determined to be too vague to evaluate. Major/field of study had the lowest rate of correct codes at 63 percent ($\chi^2 = 15.5$; $p < 0.01$).

Table 18. Summary of B&B:93/03 recode results

Type of coding	Coding attempts sampled	Percent original code correct	Percent text string too vague to code
Total	10,380	72.5	4.4
Occupation	4,620	72.2	0.2
Industry	3,720	70.9	8.2
Major/field of study	520	62.8	10.5
License/certificate	1,520	80.5	5.9

NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

A comparison of recode results by mode of data collection is presented in table 19. In general, interviewers tended to do somewhat better than sample members in selecting the correct code. Overall, the original codes assigned by interviewers were considered correct by expert coders (75 percent) more often than were codes selected by self-administered respondents (68 percent; $\chi^2 = 72.1$; $p < 0.0001$). The rates at which original codes were determined to be correct were higher for interviewer-provided responses for the occupation ($\chi^2 = 48.3$; $p < 0.0001$) and industry ($\chi^2 = 35.1$; $p < 0.0001$) coding systems than they were for respondent-provided

Table 19. Summary of B&B:93/03 recode results, by mode of interview administration

Type of coding	Web respondents			CATI/CAPI		
	Coding attempts sampled	Percent original code correct	Percent text string too vague to code	Coding attempts sampled	Percent original code correct	Percent text string too vague to code
Total	3,960	68.1	4.4	6,430	75.1	4.5
Occupation	1,790	66.6	0.4	2,830	75.8	0.1
Industry	1,310	65.9	8.1	2,410	73.6	8.3
Major/field of study	240	67.1	6.8	290	59.2	13.6
License/certificate	620	77.9	7.0	900	82.4	5.1

NOTE: CATI = Computer-assisted telephone interview; CAPI = Computer-assisted personal interview. Detail may not sum to total because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

codes. However, these rates did not differ for the major/field of study ($\chi^2 = 7.1$; $p = 0.07$) and licensing/certification ($\chi^2 = 5.0$; $p = 0.17$) coding systems.

4.2 Effectiveness of the Instrument

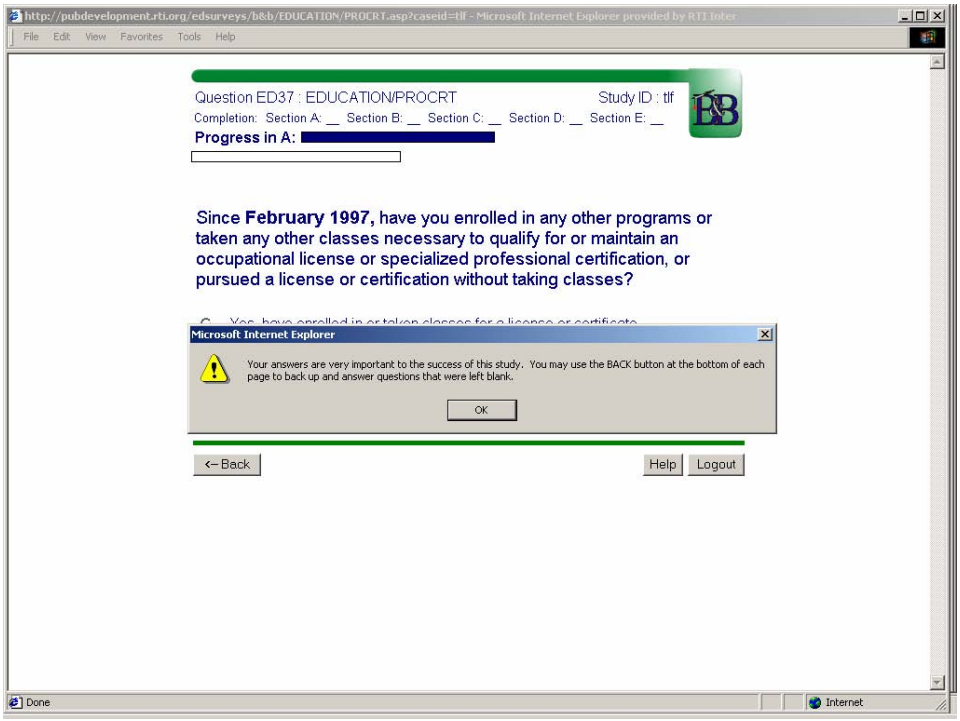
During the B&B:93/03 full-scale data collection, data quality was evaluated, in part, by the effectiveness of the web-based instrument in collecting the desired data. Evaluations of rates of indeterminate responses and of incomplete interviews (break-offs) assessed the completeness of the data collected, and a reliability reinterview assessed the temporal stability. The results of each of these evaluations are presented below.

4.2.1 Indeterminate Responses

An important measure of data quality is the rate of item-level nonresponse. Achieving low rates of item-level nonresponse is particularly important when surveys offer a self-administered component since interviewers are not present to persuade respondents to provide a definitive response. In the B&B:93/03 field test instrument, "don't know" and "refuse to answer" were provided as response options for every item. Results of the field test showed that having these explicit options may have encouraged nonresponse. Self-administered respondents were twice as likely to provide an indeterminate response as were their CATI/CAPI counterparts (Wine et al. 2004).

To minimize item-level nonresponse, several changes were made to the full-scale instrument. First, the "don't know" and "refuse" options were removed from the screen entirely. Instead, respondents could use the "continue" button to proceed without answering if an answer was unknown or they refused to answer a particular item. Second, if respondents continued through three consecutive items without providing a response, a generic pop-up box was presented to remind them of the importance of their continued participation in the interview. The pop-up box is presented in figure 5.

Figure 5. Pop-up box presented when respondents failed to respond to three consecutive questions in the B&B:93/03 full-scale interview



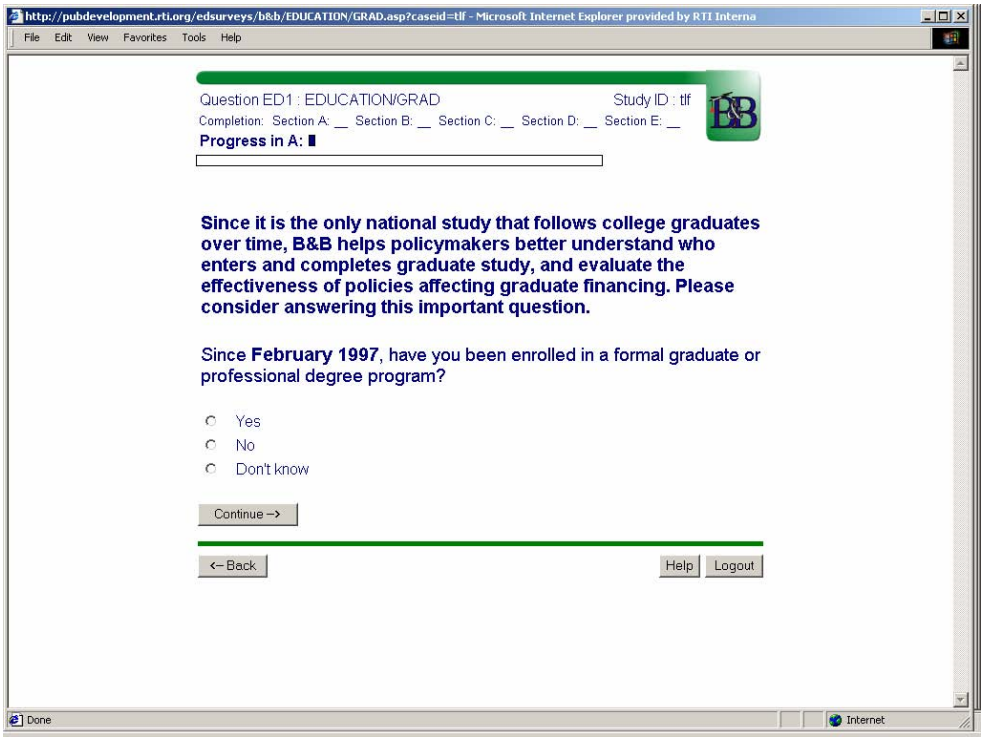
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

Another change in the full-scale interview required identifying several items as providing key information for the final follow-up. For these items, tailored text was prepared describing why a particular item was important to the study. If one of the key items was not answered, it was displayed again with special text included to encourage respondents to provide an answer. This conversion text provided respondents more specific information on why the question being asked was of particular importance to the success of the B&B:93/03.

With the presentation of the conversion text, a "don't know" option was added to the choices of response options. For the four key items on income, the conversion text was presented but, rather than providing a specific dollar value, respondents selected from among several categorical income ranges, and the "don't know" option. A sample of one of the items with conversion text is shown in figure 6. Once presented with the conversion text, respondents could select one of the original response options, choose the "don't know" option, or continue without providing a response. The effectiveness of this approach to converting indeterminate responses is discussed below.

The last modification to the full-scale interview was to change the nature of the information requested for particularly sensitive items that had high rates of missing data in the field test. For example, in the field test interview, respondents were asked to provide specific dollar amounts for any assets they held at the time of the interview. In the full-scale interview, respondents were asked more general questions about the types of savings vehicles they used. The new format decreased the rate of nonresponse for these sensitive items.

Figure 6. Sample response conversion text presented when respondents to the B&B:93/03 full-scale interview did not respond to key interview items



SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

Together, the strategies implemented for the full-scale interview served to reduce item-level indeterminacy over the field test. Only the 20 full-scale items (3 percent of the total number of items in the interview) shown in table 20 had missing data at a rate of 10 percent or higher. It is worth noting that there was no difference in the overall rate of missing data when the interview was self-administered (2 percent) and when it was interviewer-administered (2 percent; $z = 0.19$; $p > 0.10$).

For the full-scale interview, none of the items in the education section had item nonresponse at 10 percent or higher. In the employment section, only the item that asked if the respondent was looking for work while he/she was not working in February had 12 percent nonresponse. Web respondents were less likely to provide an indeterminate response (7 percent) than were CATI/CAPI respondents (20 percent, $\chi^2 = 15.0$; $p < 0.0001$).

In the teaching section, teachers were asked about the dates they began and ended teaching jobs held since 1997 and about the schools in which they taught during the 6 years elapsed between 1997 and 2003. Up to 16 percent of teachers asked could not provide information on the school that employed them prior to their current school. However, since the questions required recall of dates and other details, the observed nonresponse was likely the result of not knowing an answer rather than refusing to provide an answer. Twelve percent of teachers did not provide a response for the item asking for the primary reason they left teaching. No differences were observed in the rates of indeterminacy by mode for the items in the teacher section.

Table 20. B&B:93/03 interview item nonresponse for items with more than 10 percent missing

Interview item, by section	Number asked	Percent blank[1]
Employment		
Looked for work in February	450	11.5
Teachers		
Teaching beginning month	720	13.0
Teaching beginning year	720	15.5
Teaching school name	720	15.2
Teaching ending month	720	13.5
Teaching ending year	720	13.2
Teaching school state	660	12.5
Teaching school city	720	12.8
Teaching school identifier	720	11.6
Non-teaching position planned	500	12.2
Finances		
Year that spouse repaid loans	1,210	16.3
Spouse loans: total amount	5,780	12.3
Year repaid loans	260	10.1
Background		
Impairment: hearing	280	30.8
Impairment: visual	280	30.8
Impairment: speech	280	30.8
Impairment: mobility	280	30.8
Impairment: learning disability	280	30.8
Impairment: mental	280	30.8
Impairment: other	280	30.8

[1] Item nonresponse rates were calculated based on the number of sample members for whom the item was applicable and asked.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

Because nearly 25 percent of the field test finance questions resulted in rates of indeterminacy of 10 percent or more, many of the items were revised prior to the full-scale data collection to be less sensitive. As a result, only three items in the finance section for the full-scale interview had an indeterminacy rate of 10 percent or more. These items asked about education loan repayment, including the year the respondent's education loans were repaid, the amount of education loans owed by the spouse, and the year in which the spouse's loans were repaid. Spouse's total loan amount had an indeterminacy rate of 12 percent. Web respondents (15 percent) were more likely than CATI respondents to provide an indeterminate response for this item (9 percent, $\chi^2 = 42.5$; $p < 0.001$). Items related to the year of loan repayment for both the respondent and spouse had high rates of missing data at 10 percent and 16 percent, respectively. Web respondents were more likely to provide an indeterminate response (21 percent) for the year of spouse's loan repayment than CATI/CAPI respondents (8 percent, $\chi^2 = 36.6$; $p < 0.0001$).

As shown in table 20, only one item in the background section, which asked respondents who had previously indicated having a disability to report their disabling condition, had an indeterminacy rate of at least 10 percent. In fact, the rate of nonresponse (31 percent) was quite high, perhaps due to the sensitivity of the question. No mode differences were observed in the rates of indeterminacy for this item.

Table 21 presents the results of offering conversion text for 20 key items in the full-scale interview. For each item targeted, the number reaching the text is shown, together with the number and percentage providing a definitive response (i.e., either selecting a response option or choosing "don't know"). Also shown is the number and percentage who continued through the item without providing a response, effectively refusing to respond. The conversion text successfully converted responses from missing for 18 of the 20 items targeted. The item with the lowest conversion rate at 33 percent—currently looking for work—had only three nonrespondents. After reading the conversion text, 52 percent of respondents who initially did not provide a response to the monthly rent or mortgage payment item ended up providing a response (either definitive or don't know).

Table 21. Effectiveness of directed text in converting nonresponse to key interview items: 2003

Key interview items, by section	Number reaching conversion text	Percent of converted responses	Percent of refusal responses
Education			
Attended a formal graduate program	10	100.0	0.0
Employment			
Current employment status	40	94.6	5.4
Current/most recent job title	50	79.3	20.8
Current job: hours per week	20	75.0	25.0
Currently looking for work	#	33.3	66.7
Salary range values	650	82.8	17.2
Teacher salary range values	70	87.0	13.0
Teachers			
Ever worked as teacher or aide	20	88.9	11.1
Currently considering teaching	20	100.0	0.0
Currently employed as teacher	20	100.0	0.0
Done anything to prepare self to teach	30	100.0	0.0
Ever licensed/certified to teach	#	100.0	0.0
Finances			
Estimated income range	680	81.0	19.1
Estimated total household income range	890	74.2	25.8
Undergraduate loans: total amount	140	82.0	18.0
Undergraduate loans: amount owed	110	90.5	9.5
Postbaccalaureate loans: total amount	80	73.5	26.5
Postbaccalaureate loans: amount owed	20	100.0	0.0
Monthly payment on education loans	30	73.1	26.9
Monthly rent or mortgage payment	540	52.0	48.0

Rounds to zero.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

Responses could have been converted to one of the provided response options or to a "don't know" response. As shown in table 22, the percentage of respondents providing an explicit response (i.e., a response other than "don't know" for an item) ranged from 0 percent to 100 percent. Even for those income items considered most sensitive, response conversion was high.

Table 22. Effectiveness of directed text in evoking an explicit response to key interview items: 2003

Key interview items by section	Number converted	Percent provided explicit response
Education		
Attended a formal graduate program	10	90.9
Employment		
Current employment status	40	100.0
Current/most recent job title	40	100.0
Current job: hours per week	20	40.0
Currently looking for work	#	0.0
Salary range values	530	91.6
Teacher salary range values	60	88.3
Teachers		
Ever worked as teacher or aide	20	93.8
Currently considering teaching	20	83.3
Currently employed as teacher	20	100.0
Done anything to prepare self to teach	30	96.4
Ever licensed/certified to teach	#	100.0
Finances		
Estimated income range	550	90.1
Estimated total household income range	660	78.8
Undergraduate loans: total amount	110	56.1
Undergraduate loans: amount owed	100	50.5
Post-baccalaureate loans: total amount	60	50.8
Post-baccalaureate loans: amount owed	20	66.7
Monthly payment on education loans	20	57.9
Monthly rent or mortgage payment	280	52.9

Rounds to zero.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

4.2.2 Break-offs

Of the 8,970 interviews conducted during the full-scale B&B:93/03, only 1 percent were partial interviews (i.e., broken off before the end of the interview). To be considered a partial interview rather than a nonresponse, a respondent had to answer questions at least through the end of the first section, education. Slightly more than half of the break-offs (53 percent) occurred at some point in the employment section. There was not a particular point within a section at which the majority of break-offs occurred. In the teaching section, about one-third (30 percent) of the break-offs occurred at the first question. No other pattern of interview break-off was observed.

4.2.3 Reliability of Responses

As they completed the main B&B:93/03 interview, a subsample of 500 respondents was selected at random to complete a reinterview designed to assess the temporal stability of selected interview items. The reinterview sample was evenly divided by mode of response with 250 self-administered respondents and 250 telephone interview respondents chosen. Preloaded information and gate questions from the initial interview were preloaded for the reinterview to ensure that questions were asked in the same way and with the same wording across the two interviews. Reinterviews were conducted in the same mode as the initial interview, about 3 weeks following its completion. By the end of data collection, 36 percent of the self-administered respondents and 75 percent of the telephone interview respondents completed the interview.

Responses in the initial interview and the reinterview were compared to determine the percentage of reinterview responses that matched the original responses from the main interview. For categorical data, agreement required an exact match of interview and reinterview responses. For continuous data, responses were considered to agree when reinterview values were within one standard deviation of the main interview values. The results of the reliability reinterview analysis are presented in table 23 by interview section. Overall, percent agreement rates for the full-scale interview, which ranged from 71 to 97 percent, showed marked improvement over agreement rates achieved during the field test reinterview, which ranged from 55 to 91 percent.

Twelve items from the education section were included in the reinterview. Reliability for these items was good, with percent agreement rates ranging from 73 to 93 percent. The first set of questions asked about aspects of the respondent's undergraduate education that he/she considers very important to his/her life now. Generally, these questions had moderate percent agreement rates (between 73 and 79 percent). One item, selected if none of the listed aspects were important, had a high percent agreement rate (90 percent).

A second set of education items asked respondents to select those aspects of their life now for which they believe their undergraduate education was very important preparation. The pattern of reliability results was consistent with those of the previous set of education items, with percent agreement rates ranging from 74 to 93 percent. Again, the "none of the above" item had a high percentage of agreement (93 percent). The other two items in this section, both related to professional certification, provided reliability results with percent agreement of 77 and 83 percent.

Like education, percent agreement rates calculated for the employment section were fairly strong overall. Five questions asking respondents to rate which factors were very important to their current/most recent employment, like similar items in the education section, showed the greatest range in agreement rates (71 percent to 97 percent). The question asking respondents to confirm employment in February of 2003 had high agreement from initial interview to reinterview (95 percent). Of the mismatches in response between interviews, 4.4 percent resulted from respondents changing a "no" response to "no-for a specific reason" response, or the converse.

Eight questions from the finance section were included in the reinterview. As discussed above in the section on indeterminate responses, these items were redesigned after the field test data collection to be less sensitive with the goal of improving item-level response rates. Agreement for these items ranged from 77 to 92 percent.

Table 23. Percent agreement for items in the reinterview, by main interview section: 2003

Variable label	Percent agreement[1]
Education	
Value of undergraduate education: particular major(s) chosen	76.0
Value of undergraduate education: liberal arts courses taken	79.1
Value of undergraduate education: professional courses taken	72.5
Value of undergraduate education: quality of instruction	76.3
Value of undergraduate education: internship and other work	75.3*
Value of undergraduate education: none of the above	90.3
Undergraduate preparation: work and career	85.9
Undergraduate preparation: further education	74.4
Undergraduate preparation: financial security	80.3
Undergraduate preparation: none of the above	92.5
Graduate/undergraduate program required for certification	82.9
Took classes for professional certification	76.6
Employment	
More than one career in last 10 years	83.2
Expect same type of work in 3 years	91.2
Working in February 2003	94.7
Importance of undergraduate education	75.6*
Importance of graduate education	97.0
Importance of any other education	73.1
Importance of formal on the job training	71.2
Importance of experience from other jobs	78.8
Taken any leave for children	78.8
Current job: year started	95.2
Finances	
Own collectibles	87.6
Savings: savings account	84.4
Savings: money market	82.5
Savings: certificate of deposit	88.4
Savings: stocks/bonds/mutual funds	84.1
Savings: retirement account	92.2
Savings: cash value life insurance	76.9
Savings: other	84.4

* Difference between web respondents and CATI respondents statistically significant ($p < 0.05$).
[1] Percent agreement reflects an exact match of the paired interview/reinterview responses for categorical items.
NOTE: Analyses were conducted only for respondents with determinate response on both the initial interview and the reinterview; not all questions applicable to all respondents.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

As part of the reinterview analysis, percent agreement rates for self-administered and telephone interview respondents were compared to determine the extent of difference in rates due to mode of interview completion. Two items in the reinterview—value of undergraduate internship and other work experience and importance of undergraduate education to current employment—showed statistically significant differences by mode. The first item, on internships and other work, had a statistically significant difference in the rate of agreement by mode ($\chi^2 = 4.25$; $p < 0.05$) such that self-administered responses showed higher percent agreement over time (81 percent) than did those who completed a telephone interview (71 percent). For the item rating the importance of undergraduate education, percent agreement was higher for those completing the telephone interview (80 percent) than for those who completed the self-administered interview (70 percent; $\chi^2 = 4.33$; $p < 0.05$). No other differences by mode were observed.

In summary, percent agreement for items included in the reinterview was good overall, suggesting that responses were reliable over time. The somewhat mixed rates for reinterview items requiring respondents to indicate if something was "very important" were not surprising for two reasons. First, while reinterviews generally occurred as early as 3 weeks after the initial interview, many were actually conducted much later than that—up to 3 months later—due to delays locating respondents and gaining their continued cooperation. Respondents' value ratings could have changed naturally in the time between the first and second interviews. In addition, since these items required respondents to define "very important" for themselves, this subjectivity may have introduced additional response variance between the initial interview and reinterview.

4.3 Effectiveness of the Data Collection Design

Effectiveness of the data collection design for the B&B:93/03 field test, the final measure of data quality, was measured through quality assurance monitoring and quality circle meetings. Results for both evaluations are presented below, along with a summary of quality circle meetings held throughout data collection.

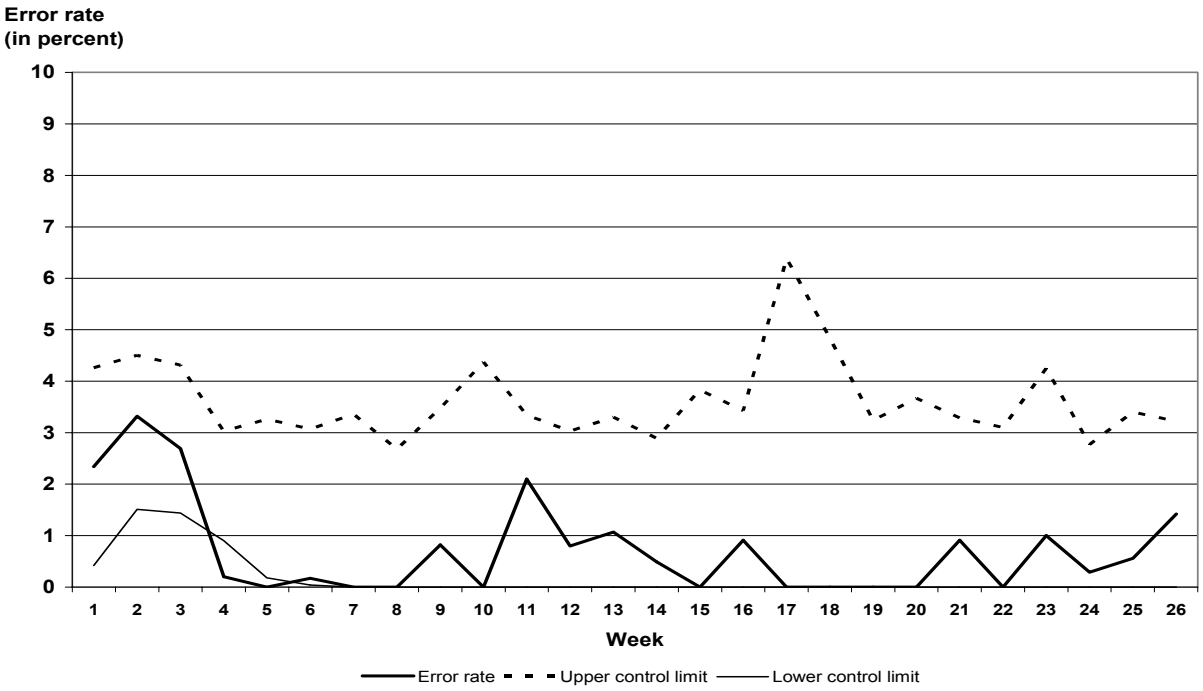
4.3.1 Quality assurance CATI monitoring

Regular monitoring of telephone interviews leads to better interviewing and data quality as well as improvements in data collection costs and in the efficiency of the telephone facilities. To ensure that sufficient monitoring occurred for the full-scale B&B:93/03, monitoring sessions were conducted during day, evening, and weekend shifts. Monitors listened to and simultaneously viewed the progress of interviews using remote monitoring telephone and computer equipment. Monitors listened to up to 20 questions during an ongoing interview and, for each question, evaluated two aspects of interviewer performance: (1) correct delivery of questions (error in delivery) and (2) accurate keying of the response (error in data entry).

Measures of question delivery and data entry were developed and daily, weekly, and cumulative reports produced. Monitoring took place throughout data collection, with a total of 10,640 items monitored. During the final weeks of data collection, monitoring efforts were scaled back due to lighter caseloads. Among the 10,640 items observed, only 115 delivery errors and 66 data entry errors were observed. Error rates in delivery and data entry, by week of data

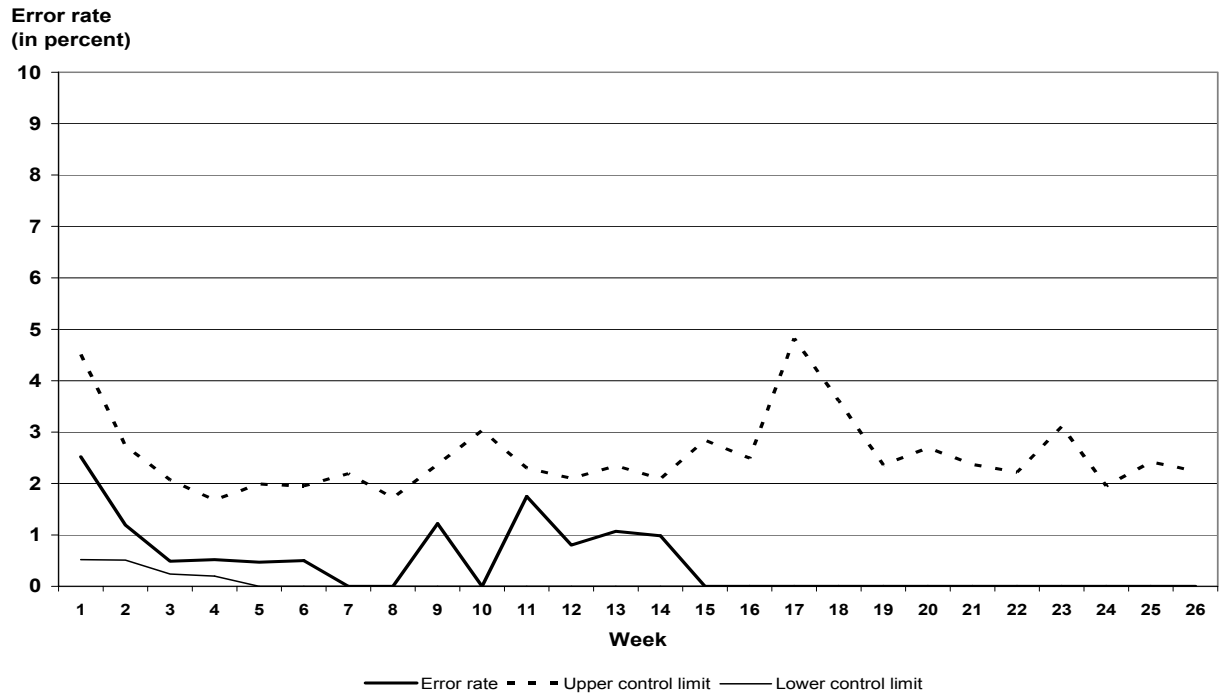
collection are shown in figures 7 and 8, respectively. The relatively high error rates in the early weeks and in Week 11 of data collection can be attributed to the assignment of newly trained interviewers to the study. The spikes in the upper control limit are due to the low number of monitoring observations during that period.

Figure 7. Error rates for computer-assisted telephone interviewing (CATI) question delivery: 2003



NOTE: The upper and lower control limits were defined by three times the standard error of the proportion of errors to the number of questions observed for the period (upper control limit : + 3 times the standard error; lower control limit: -3 times the standard error).
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

Figure 8. Error rates for computer-assisted telephone interviewing (CATI) data entry: 2003



NOTE: The upper and lower control limits were defined by three times the standard error of the proportion of errors to the number of questions observed for the period (upper control limit : + 3 times the standard error; lower control limit: -3 times the standard error).
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

4.3.2 Quality circle meetings

Quality circle meetings provided an opportunity for B&B:93/03 interviewers to discuss data collection issues with project staff. Topics discussed during these meetings covered all aspects of data collection, including Help Desk, tracing and locating, and interviewing. Meetings were scheduled weekly during the day and evening shifts to ensure that all telephone interviewers had an opportunity to attend. Summaries of the discussions and decisions addressed during these meetings were compiled and distributed to all interviewers in the form of a newsletter. Issues covered in quality circle meetings included problem sheets, coding strategies, achieving gatekeeper cooperation, interview logic, and clarification of the intent of questions and help text.

This page left intentionally blank.

Chapter 5
Variable Construction and File Development

As the fourth and final interview with the B&B:93 cohort, the data files for the 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03) contain a number of component data files from a variety of sources in addition to those files created from the interview itself. These files are available as a set of restricted research files, fully documented by an electronic codebook (ECB), and as a public release Data Analysis System (DAS), which also contains full documentation.[9] This chapter describes each data file and details the data editing and documentation process.

5.1 Overview of the B&B:93/03 Data Files

The B&B:93/03 data files contain student-level and institution-level data collected from student interviews and government financial aid databases. The primary analysis file, from which the study DASs were constructed, contains data for approximately 11,100 study respondents. The primary analysis file contains over 2,900 variables, developed from B&B:93/03 interview data and data from previous interviews with the B&B:93 cohort. Throughout the data collection period, data were processed and examined for quality control purposes. Editing of student data began shortly after the start of self-administered web data collection, when procedures and programs for this purpose were first developed. Anomalous values were investigated and resolved, where appropriate, through the use of data corrections and logical imputations.

Complete data for B&B:93/03 are located on the restricted access files and are documented by the ECB. The restricted files and the ECB are available to researchers who have applied for and received authorization from NCES to access restricted research files. Authorization may be obtained by contacting the NCES Data Security Office.

The restricted use B&B:93/03 ECB contains information about the following files:

- *2003 Derived Variables File*—Contains the composite (derived) variables developed for use on the B&B:93/03 public release DAS. [B03DAS.DAT]
- *2003 Student Interview Data File*—Provides student-level raw data collected from the 9,000 sample members who responded to the B&B:93/03 interview. The file excludes any "verbatim" variables, which are on the Verbatim Data File described below. [B03STUD.DAT]
- *2003 Postsecondary Institution Data File*—Provides data obtained from the student interview on postsecondary institutions attended since the 1997 interview (B&B:93/97). Although this is a student-level file, a sample member may have more than one record in the file since there is a separate record for each postsecondary institution attended by the sample member. [B03SCH.DAT]

[9] The ECB and DAS are both fully documented software products available from the National Center for Education Statistics (NCES). THE DAS is available online at http://nces.ed.gov/das.

- *2003 Elementary/Secondary School Data File*—Provides school data obtained from the student interview for all elementary and secondary schools in which a respondent taught since the last interview. Records were created from the most recent Common Core of Data (CCD) and Private School Survey (PSS) databases available from NCES. [PSS.DAT, CCD.DAT, UNCODSCH.DAT]
- *2003 Coding Results File*—Contains the verbatim text strings and resulting codes from the coding systems used during the 2003 student interview. For respondents who have pursued additional postsecondary education, major field of study is provided. For employed students, industry and occupation are included. One record is provided per student. [B03CODE.DAT]
- *2003 Interview Preload File*—Contains those data preloaded from earlier data collections for use during the 2003 interview for the 8,970 respondents. Some preloaded variable values were updated as a result of the interview. Consequently, caution is needed when using this file for analytic purposes. [B03PREL.DAT]
- *CPS Data Files*—Contains data received from matches to the Central Processing System (CPS) [10] database for each consecutive academic year (AY) since the last follow-up. From 1999–2000 through 2002–03, up to 420 sample member records were matched to the CPS. [CPS9900.DAT, CPS0001.DAT, CPS0102.DAT, CPS0203.DAT]
- *NSLDS Loan File*—Contains raw loan-level data from the National Student Loan Data System (NSLDS) for the students who received federal loans originated at any time before December 2003. This is a history file with multiple records for each student. [NSLDS.DAT]
- *Weights File*—Contains one analysis weight and 42 Balanced Repeated Replication (BRR) weights created for cross-sectional analysis of B&B:93/03. In addition, a panel weight and 42 BRR weights are included for longitudinal analysis of those who responded to each of the four interviews—base year, B&B:93/94, B&B:93/97, and B&B:93/03. Weights for each of the earlier interviews are included on their respective files (described below). [B03WEIGHT.DAT]
- *NPSAS:93 File*—Contains the base-year data included in the NPSAS:93 ECB subset to the B&B:93 sample. Includes the derived variables contained on the DAS, the derived CADE and CATI variables, the weights file and the derived variables from the National Student Loan Data System (NSLDS). [N93B03.DAT; N93BDERV.DAT; N93BCADE.DAT; N93BCATI.DAT; N93BRTI.DAT; N93BB4NS.DAT]
- *B&B:93/94 File*—Contains all data on the B&B:93/94 first follow-up ECB. Includes derived DAS and CATI variables, the weights file, and files created from the student transcript abstraction. [BB94ECB6.DAT; SCGRADE.DAT; SCGRADE.TXT; STUDMST.DAT; STUDCRS.DAT; SCHLMST.DAT]

[10] The Central Processing System is a database maintained by the Department of Education which contains data from the Free Application for Federal Student Aid (FAFSA) for all students who applied for federal aid.

- *B&B:93/97 File*—Contains all data on the B&B:93/97 second follow-up. Includes derived DAS and CATI variables and the weights file. [B97DAS.DAT; B97ECB7.DAT]

5.2 Data Coding and Editing

The B&B:93/03 data were coded and edited using procedures developed and implemented for previous National Center for Education Statistics (NCES)-sponsored studies. The coding and editing procedures fell into two categories: online coding and editing performed during data collection, and post-data-collection data editing.

5.2.1 Online Coding and Editing

The B&B:93/03 follow-up study used one major system—a web instrument—for all data collection. The web instrument included online coding systems which categorized user-provided input into specific codes for major field of study, occupation, and industry. In addition, online coding systems were used to collect Integrated Postsecondary Education Data System (IPEDS) information for all postsecondary institutions attended since the B&B:93/97 interview and elementary/secondary school information for all respondents who taught at the elementary and/or secondary level since the last interview.

The web-based data collection system also included edit checks to ensure that the data collected were within valid ranges. To the extent feasible, this system incorporated across-item consistency edits. Whereas more extensive consistency checks would have been technically possible, use of such edits was limited to prevent excessive respondent burden. Below is a description of the online range and consistency checks incorporated into the B&B:93/03 web instrument.

General Verifications

- Range checks were applied to all numerical entries, such that only valid numeric responses could be entered.
- If, in response to a "check all that apply" question, a valid answer and the "none of the above" option were both checked, respondents and interviewers were advised to uncheck other options before checking the "none of the above" option.
- Pop-up messages confirmed responses which fell outside prespecified ranges for selected numeric values such as income and hours worked per week.
- Consistency checks identified conflicting responses (e.g., if the beginning date for a job was later than the end date provided, or if the highest grade taught was lower than the lowest grade taught) and allowed respondents the opportunity to change answers as appropriate.

5.2.2 Post-Data-Collection Editing

Both during and upon completion of data collection, edit checks were performed on the B&B:93/03 data file to confirm that the intended skip patterns were implemented during the

interview. At the conclusion of data collection, special codes were added as needed to indicate the reason for missing data. Missing data within individual data elements can occur for a variety of reasons. Table 24 lists each missing value code and its associated meaning in the B&B:93/03 interview.

Table 24. Description of missing data codes: 2003

Missing data code	Description
–1	Don't know
–3	Not applicable
–6	Bad data, out of range
–7	Item was not reached (partial interviews)
–8	Item was not reached due to an error
–9	Data missing, reason unknown

SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

Skip-pattern relationships in the database were examined by methodically running cross-tabulations between gate items and their associated nested items. In many instances, gate-nest relationships had multiple levels within the instrument. That is, items nested within a gate question may themselves have been gate items for additional items (e.g., citizenship serves as a date for voter registration which serves as a gate for voting behavior items). Therefore, validating the gate-nest relationships often required much iteration and many multiway cross-tabulations.

The data editing process for the B&B:93/03 data involved a multistage process that consisted of the following steps:

Step 1. Blank or missing data were replaced with -9 for all variables in the instrument database. A one-way frequency distribution of every variable was reviewed to confirm that no missing or blank values remained. These same one-way frequencies revealed any out-of-range or outlier values, which were investigated and checked for reasonableness against other data values. Example: hourly wages of $0.10, rather than $10.00. Creating SAS formats from expected values and the associated value labels also revealed any categorical outliers.

Descriptive statistics were produced for all continuous variables. All values less than zero were temporarily recoded to missing. Minimum, median, maximum, and mean values were examined to assess reasonableness of responses and anomalous data patterns were investigated and corrected as necessary.

Step 2. Legitimate skips were identified using the interview source code. Gate-nest relationships were defined to replace -9's (missing for unknown reason) with -3's (not applicable) as appropriate. Two-way cross-tabulations between each gate-nest combination were evaluated, and high numbers of nonreplaced -9 codes were investigated to ensure skip-pattern integrity.

Nested values were further quality checked to reveal instances in which the legitimate skip code over-wrote valid data which typically occurred if a respondent answered a gate question and the appropriate nested item(s), but then backed up and changed the value of the gate, following an alternate path of nested item(s). Responses to the first nested item(s) remained in the database and, therefore, required editing.

In cases where it could not be determined whether nested items had been legitimately skipped because the response to the gate item was indeterminate (either blank, -9, or don't know, -1), the edit code replaced -9's in nested items with the same value as the gate item. In this way, the value of the gate item was carried through to the nested items.

Step 3. Variable formatting (e.g., formatting dates as YYYYMM) and standardization of time units, for items which collected amount of time in multiple units, were performed during this step. In addition, any new codes assigned by expert coders reviewing major field of study, occupation, industry, IPEDS, and elementary/secondary school codes first selected through the online coding systems were merged back with the interview data files.

Also at this step, logical imputations were performed when the value of missing items could be determined from answers to previous questions or preloaded values. For example, if the respondent indicated that he/she had not worked since 1997, then the number of different jobs and the number of employers since 1997 were logically imputed to "0" rather than -3 or -9 even though the questions were skipped in the interview.

Items skipped because preload data already contained a valid value were also imputed. For example, if undergraduate loan amount was available as a preload, that question was skipped. Undergraduate loan amount was imputed from prior interviews for these cases.

Step 4. During the interview, postsecondary institutions were coded for all respondents who enrolled in a formal degree program since the last interview using the IPEDS database. For respondents who were teachers at the K-12 level, the elementary/secondary schools in which they taught were also coded using the CCD database on public elementary and secondary education and the PSS database. Following data collection, these files were merged by the school code to obtain additional information including level, control, district, county, etc. for delivery with the B&B:93/03 data.

Step 5. At this step, special codes of -3 and -9 were replaced with -7 (item not administered) based on the section completion indicators. The -7 code allows analysts to easily distinguish items not administered from items that were either skipped or simply left blank.

Step 6. One-way frequency distributions for all categorical variables and descriptive statistics for all continuous variables were examined. Out-of-range or outlier values were either replaced with the value of -6 (bad data, out of range) or recoded to a more reasonable value. For example, if a respondent reported income of more than $500,000, that income value was set to $500,000, the maximum amount allowed for the item.

Step 7. One-way frequencies on all categorical variables were regenerated and examined. Variables with high counts of -9 values were investigated. Because self-administered web respondents could skip over most items without providing an answer, -9's did remain a valid value, especially for sensitive items, such as those asking for financial information.

Concurrent with the data editing process, detailed documentation was developed to describe question text, response options, logical imputations, recoding, and the "applies to" text for each delivered variable.

5.3 Composite and Derived Variable Construction

Analytic variables were created by examining the data available for each student from the various sections of the survey instrument. In some cases, raw interview items were recoded or otherwise summarized to create derived variables. In other cases, information from two or more survey items was combined to create a third, composite variable. A listing of the set of analysis variables derived for B&B:93/03 appears in appendix F. Specific details regarding the creation of each variable appear in the variable descriptions contained in the ECB and DAS.

Chapter 6
Weighting, Variance Estimation, and Imputation Methodology

Development of statistical analysis weights for the 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03) sample is discussed in section 6.1 below. Cross-sectional weights were constructed for analyzing the 8,970 respondents to B&B:93/03. In addition, a panel (longitudinal) weight was constructed for analyzing the 8,090 students who responded to all four surveys: B&B:93/03, B&B:93/97, B&B:93/94, and 1992–93 National Postsecondary Student Aid Study (NPSAS:93).

Analysis procedures that can be used to produce design-unbiased estimates of sampling variances are discussed in section 6.2, including variances computed using Taylor series and balanced repeated replications (BRR) techniques. Section 6.2 also describes how the Taylor series strata and primary sampling unit (PSU) variables were constructed, and how the BRR weights were constructed.

Section 6.3 discusses the accuracy of B&B:93/03 estimates in terms of both precision and the potential for nonresponse bias. Survey design effect tables that illustrate the level of precision achieved by the B&B:93/03 survey for key analytic outcomes for several important analysis domains are included in appendix G.

Finally, section 6.4 gives the weighted response rates and an analysis of item nonresponse bias.

6.1 Analysis Weights

The initial file used for the B&B:93/03 sample frame contained 11,200 students. Of these,

- 10,090 were B&B:93/97 respondents;
- 1,070 were B&B:93/97 nonrespondents; and
- 10 were identified as deceased during the B&B:93/03 advance tracing death search.

With deceased cases excluded, a total of 11,150 students were determined to be eligible for B&B:93/03. As noted in chapter 2, the final B&B:93/03 sample consisted of the 10,090 eligible respondents to B&B:93/97 and a subsample of 360 of the nonrespondents to B&B:93/97, for a total of 10,440 sample members. During data collection for B&B:93/03, about 40 sample members were determined to be ineligible either because they were deceased or because they were determined to have not earned a baccalaureate degree during the NPSAS year (July 1, 1992 through June 30,1993), leaving 10,400 eligible sample members. Of these 10,400, a total of 8,970 responded to B&B:93/03.

A statistical analysis weight was computed to use for analyzing data from the 8,970 eligible respondents[11] to the B&B:93/03 survey. In addition, a panel weight was computed for analyzing the 8,390 respondents who participated in all four studies: B&B:93/03, B&B:93/97, B&B:93/94, and NPSAS:93.

The weights for the B&B:93/03 respondents were constructed by applying a series of adjustments for subsampling and nonresponse to the B&B:93/94 base weight (BNBWT0). Specifically, four adjustments were made to account for

- subsampling of the B&B:93/97 nonrespondents;
- those not located;
- refusals among those who were located; and
- types of nonresponse other than refusals among those who were located and did not refuse.[12]

These last three adjustments are consistent with the procedures used to adjust weights for other longitudinal postsecondary studies (e.g., 1996/01 Beginning Postsecondary Student Longitudinal Study [BPS:96/01] and B&B:2000/01).

Construction of the panel (or longitudinal) weight to be used for analyzing those who responded to all three surveys consisted of an additional adjustment for nonresponse for the B&B:93/03 respondents who did not respond to all three of the previous surveys.

6.1.1 Base Weight for B&B:93/03—Adjustment for Subsampling of B&B:93/97

As discussed in chapter 2, a subsample of 360 B&B:93/97 nonrespondents was included in B&B:93/03, rather than all nonrespondents, to reduce data collection costs. The subsample was selected using probability proportionate to size (PPS) sampling, with the size measure being the B&B:93/94 base weight (BNBWT0), after stratifying the B&B:93/97 nonrespondents by the B&B:93/03 advance tracing status, the control of the base-year school, and the B&B:93/94 response status. The base weight was adjusted for those students, *j*, in the subsample by multiplying by the inverse of their selection probabilities. These probabilities take into account the stratification and PPS sampling that was used in selecting the subsample. The adjustment was

$ADJ1_j = 1/\pi_j$, where π_j is the selection probability.

The weight was calculated as

BB03_W1U = BNBWT0 * ADJ1, for students in the B&B:93/97 nonrespondent subsample

[11] As discussed, sample members ineligible for the B&B cohort were those determined to have not earned a baccalaureate degree during the NPSAS year (July 1, 1992 through June 30, 1993).
[12] Sample members determined to be ineligible during data collection were not included in the weight adjustments for those who were not located, refused, or did not participate for some other reason. However, ineligibles were included in the weight adjustment for subsampling.

= BNBWT0 for all other students.

The weights BB03_W1U for the students in the subsample were also adjusted so that they summed to the weight sum of B&B:93/97 for the B&B:93/97 nonrespondents within categories of the control of the base-year school. This adjustment resulted in the initial sampling weight for the B&B:93/03 sample, which is denoted BB03_W1. BB03_W1 was further adjusted to produce the B&B:93/03 analysis weights, as described below.

6.1.2 B&B:93/03 Cross-Sectional Weights

Analysis weights were constructed for the 8,970 respondents to B&B:93/03. The weights were constructed by applying adjustments to the base weight BB03_W1. This section describes each of the adjustment steps and the variables used in making the adjustments.

The adjustment for nonresponse was performed in three steps because the predictors of response propensity are potentially different for each of the following outcomes:

- inability to locate the student;
- refusal to be interviewed; and
- other noninterview.

Using these three steps of nonresponse adjustment can achieve greater reduction in nonresponse bias to the extent that different variables are significant predictors of nonresponse propensity at each step. Also, as noted earlier, this is consistent with the steps used in the construction of weights for other longitudinal postsecondary studies (e.g., BPS:96/01 and B&B:2000/01).

All nonresponse adjustments were fitted using RTI's proprietary generalized exponential modeling procedure (GEM) (Folsom and Singh 2000), which is similar to logistic modeling using bounds for adjustment factors. A key feature and advantage of the GEM software is that the nonresponse adjustment and weight trimming and smoothing are all accomplished in one step. Lower and upper bounds are set on the weight adjustment factors. The bounds can be varied, depending on whether the weight falls inside or outside a range, such as one defined by the bounds (median – 3 times the interquartile range, median + 3 times the interquartile range). This allows different bounds to be set for adjustments for weights that are considered high extreme, low extreme, or nonextreme. In this way, the extreme weights can be controlled and the design effect due to unequal weighting reduced.

Candidate predictor variables were those thought to be predictive of nonresponse and nonmissing for most of the sample (nonrespondents as well as respondents). Candidate predictor variables for the B&B:93/03 weight adjustment included

- age (categorical);
- race/ethnicity;
- gender;
- citizenship status in the base year;

- attendance status in the base year;
- control of institution attended in the base year;
- region of institution attended in the base year;
- size of institution attended in the base year (categorical);
- 16-level B&B base-year institutional strata;
- student type in base year (business major, nonbusiness major, other);
- applied for financial aid in the base year (yes or no);
- receipt of federal aid in the base year (yes or no);
- receipt of Pell Grant in the base year (yes or no);
- receipt of Stafford Loan in the base year (yes or no);
- receipt of state aid in the base year (yes or no);
- receipt of institutional aid in the base year (yes or no);
- receipt of any aid in the base year (yes or no);
- previous response status (whether the student was a respondent to either B&B:93/94 and B&B:93/97 versus a nonrespondent to both B&B:93/94 and B&B:93/97);
- income of independent students and parents of dependent students (collapsed);
- number of telephone numbers available;
- number of times an answering machine was encountered (for located students); and
- whether the student was located in a field cluster.

To detect important interactions for the logistic models, a Chi-squared automatic interaction detection analysis (CHAID) was performed on the predictor variables. The CHAID analysis divided the data into segments that differed with respect to the response variable (located, did not refuse, or respondent, depending on the model). The segmentation process first divided the sample into groups based on categories of the most significant predictor of response. It then split each of these groups into smaller subgroups based on other predictor variables. It also merged categories of a variable that are found to be nonsignificant. This splitting and merging process continued until no more statistically significant predictors were found (or until some other stopping rule was met). The interactions from the final CHAID segments were then defined.

The nonresponse bias for these same variables was estimated, and then a statistical test of whether or not the bias was significant was performed. Tests were performed to identify significant differences between refusal conversions and other respondents; significant differences suggest a potential for nonresponse bias because of the refusal population being different from the other respondents. Additional tests were performed to detect significant differences between late respondents and other respondents; significant differences suggested a potential for

nonresponse bias because of the noncontacts/late-contact population being different from the other respondents. Details of the nonresponse bias analyses are given in section 6.3.

The interaction segments and all the main effects were subjected to variable screening in the GEM logistic procedure. Variables with significant bias were included in each nonresponse model. The initial models for each adjustment step included all of the potentially important variables. The interaction segments identified by CHAID were also retained in all of the models. The most insignificant variables were deleted sequentially until the deletion of additional variables did not appreciably improve the unequal weighting effect (UWE). Different bounds on the weight adjustments, depending on whether the weight is classified as high extreme, nonextreme, or low extreme, were used to accomplish nonresponse adjustment, truncation, and smoothing in one step. The UWEs did not change very much when insignificant variables were dropped, and, as a result, a large number of predictor variables were retained in each of the nonresponse model adjustments. This allows the estimates to be calibrated based on the respondents to as many totals as possible that are known for both respondents and nonrespondents.

6.1.2.1 Weight adjustment for nonrespondents who were not located

Of the 10,400 individuals determined to be eligible, 9,730 were located. An adjustment was first performed to the weight BB03_W1 to adjust for the 670 eligible sample members who did not respond because they were not located. A CHAID analysis was performed on all of the predictor variables to detect important interactions. All potentially important variables were included in the model. There was very little change in the UWE when highly insignificant variables were deleted from the model, and, as a result, all variables were retained in the model.

Table 25 presents the predictor variables used in the logistic model that adjusted the weights for those who were not located and gives the weighted location rate and the average weight adjustment factors resulting from these variables. The table includes all levels of the variables used in the model although, due to dependencies and small sample sizes, not all of the levels were used as predictors. For example, the private, not-for-profit and private, for-profit categories for institution control were included in the model, but the school enrollment category for private, for-profit institutions was not due to small cell sizes. Similarly, the public category for institution control was not included since it is a linear combination of the school enrollment categories for public institutions.

The weighting adjustment factor for student j is the reciprocal of the predicted response probability, or

$$ADJ2_j = 1/ p_{L,j}.$$

The weight, adjusted for those who were not located, was computed as

LOCWT = BB03_W1 * AJD2 for the 9,730 who were located
= 0 otherwise.

Table 25. Average weight adjustment factors from the logistic model used to adjust for student location nonresponse: 2003

Predictor variables	Number located	Weighted location rate	Average weight adjustment
Total	9,730	92.6	1.07
Age			
21 or younger	2,600	93.8	1.06
22	2,630	92.2	1.08
23	1,330	92.7	1.07
24 to 27	1,410	91.5	1.09
28 or older	1,730	92.5	1.08
Race/ethnicity			
White, non-Hispanic	8,290	94.2	1.06
Black or African American, non-Hispanic	550	84.0	1.18
Hispanic	470	88.7	1.09
Asian/Native Hawaiian/Other Pacific Islander	340	79.4	1.26
American Indian/Alaska Native	60	94.3	1.06
Gender			
Male	4,170	91.8	1.08
Female	5,560	93.3	1.07
U.S. citizenship			
Yes	9,600	93.1	1.07
No	120	73.2	1.39
Attendance status			
Full time/full year: 1 institution	4,470	93.0	1.07
Full time/full year: more than 1 institution	210	87.2	1.14
Full time/part year	1,710	92.3	1.08
Part time/full year: 1 institution	1,460	92.9	1.07
Part time/full year: more than 1 institution	140	93.3	1.08
Part time/part year	1,690	92.1	1.08
Institution control			
Public	6,330	92.7	1.07
Private, not-for-profit	3,320	92.5	1.08
Private, for-profit	80	87.3	1.13
Institution region			
New England	700	93.2	1.07
Mid East	1,630	91.2	1.09
Great Lakes	1,600	94.3	1.06
Plains	850	97.0	1.02
Southeast	2,380	91.0	1.09
Southwest	1,070	93.0	1.07
Rocky Mountains	330	92.8	1.07
Far West	1,060	91.0	1.09
Outlying areas	80	92.4	1.06

See notes at end of table.

Table 25. Average weight adjustment factors from the logistic model used to adjust for student location nonresponse: 2003—Continued

Predictor variables	Number located	Weighted location rate	Average weight adjustment
Type of institution and enrollment category			
Public			
Fewer than 2,500	150	92.6	1.07
2,500–4,999	450	91.1	1.09
5,000–9,999	990	94.3	1.06
10,000–19,999	1,500	92.7	1.07
20,000 or more	3,220	92.5	1.08
Private, not-for-profit			
Fewer than 1,000	230	89.4	1.11
1,000–2,499	1,040	93.0	1.07
2,500–4,999	770	93.2	1.07
5,000–9,999	630	93.0	1.07
10,000 or more	640	91.7	1.09
Private, for-profit			
Fewer than 1,000	40	90.4	1.04
1,000 or more	40	81.9	1.22
B&B institution stratum[1]			
Public 4-year first-professional high education	390	94.6	1.05
Public 4-year first-professional low education	2,400	92.6	1.07
Private 4-year first-professional high education	730	90.5	1.10
Private 4-year first-professional low education	270	89.7	1.11
Public 4-year doctor's high education	380	93.8	1.06
Public 4-year doctor's low education	820	92.7	1.07
Private 4-year doctor's high education	130	90.7	1.11
Private 4-year doctor's low education	140	94.1	1.06
Public 4-year master's high education	330	94.9	1.05
Public 4-year master's low education	1,760	91.4	1.09
Private 4-year master's high education	150	90.5	1.10
Private 4-year master's low education	1,110	93.8	1.06
Public 4-year bachelor's high education	110	97.1	1.03
Public 4-year bachelor's low education	130	96.6	1.03
Private 4-year bachelor's high education	110	96.4	1.04
Private 4-year bachelor's low education	760	92.9	1.06
B&B student stratum			
Other students: Combined cell	530	92.9	1.08
Business majors	800	92.4	1.08
Nonbusiness majors	8,400	92.6	1.07
Applied for aid			
Yes	5,470	92.9	1.07
No	4,020	92.4	1.08
Receipt of federal aid			
Yes	3,870	92.1	1.08
No	5,840	92.8	1.07

See notes at end of table.

Table 25. Average weight adjustment factors from the logistic model used to adjust for student location nonresponse: 2003—Continued

Predictor variables	Number located	Weighted location rate	Average weight adjustment
Receipt of Pell Grant			
Yes	2,190	91.6	1.09
No	7,520	92.8	1.07
Receipt of Stafford Loan			
Yes	2,960	92.1	1.08
No	6,750	92.8	1.07
Receipt of state aid			
Yes	1,500	92.9	1.07
No	8,200	92.6	1.08
Receipt of institution aid			
Yes	2,500	93.9	1.07
No	7,200	92.3	1.08
Receipt of any aid			
Yes	5,270	92.9	1.07
No	4,430	92.3	1.08
Prior respondent			
Respondent to either B&B:93/94 or B&B:93/97	9,560	93.4	1.07
Nonrespondent to both surveys	170	78.9	1.27
Parents' income (for dependent students)			
Less than $10,000	200	93.1	1.07
10,000–19,999	360	91.0	1.10
20,000–29,999	550	92.8	1.07
30,000–39,999	600	92.9	1.07
40,000–49,999	710	93.5	1.06
50,000–59,999	1,010	91.8	1.08
60,000–69,999	720	92.4	1.07
70,000–79,999	450	93.9	1.06
80,000–99,999	510	92.4	1.09
100,000 or more	690	94.0	1.06
Student's income (for independent students)			
Less than $5,000	750	90.6	1.09
5,000–9,999	740	91.6	1.09
10,000–19,999	800	91.0	1.09
20,000–29,999	480	91.5	1.08
30,000–49,999	630	95.4	1.03
50,000 or more	330	94.7	1.06
Telephone numbers available			
0 or 1	450	88.4	1.12
2	1,360	92.2	1.08
3	2,290	93.9	1.06

See notes at end of table.

Table 25. Average weight adjustment factors from the logistic model used to adjust for student location nonresponse: 2003—Continued

Predictor variables	Number located	Weighted location rate	Average weight adjustment
Telephone numbers available—Continued			
4	2,270	93.2	1.07
5	2,100	93.3	1.07
6	860	91.2	1.10
7 or more	390	88.9	1.13
Number times answering machine encountered			
0	2,970	94.2	1.05
1	910	95.7	1.04
More than 1	5,850	91.4	1.09
In field cluster			
Yes	4,390	92.3	1.07
No	5,330	92.9	1.07
Interaction segment			
Did not respond to B&B:93/94 or B&B:93/97	170	78.9	1.27
Prior respondent, White or American Indian/Alaskan Native, answering machine 0 or 1 time, 0 or 1 telephone number	210	89.4	1.12
Prior respondent, White or American Indian/Alaskan Native, answering machine 0 or 1 time, 2 telephone numbers	610	94.9	1.06
Prior respondent, White or American Indian/Alaskan Native, answering machine 0 or 1 time, 3 or more telephone numbers	2,500	97.9	1.02
Prior respondent, White or American Indian/Alaskan Native, answering machine 2 or more times, 5 or fewer telephone numbers	3,980	94.0	1.06
Prior respondent, White or American Indian/Alaskan Native, answering machine 2 or more times, 6 or more telephone numbers	910	91.2	1.10
Prior respondents, Black/Hispanic/Asian/Hawaiian/ Other, U.S. citizen, Dependent income less than $30,000	250	87.9	1.14
Prior respondent, Black/Hispanic/Asian/Hawaiian/ Other, U.S. citizen, Dependent income $30,000–49,999	190	90.8	1.10
Prior respondent, Black/Hispanic/Asian/Hawaiian/ Other, U.S. citizen, Dependent income $50,000–59,999	90	86.3	1.17
Prior respondent, Black/Hispanic/Asian/Hawaiian/ Other, U.S. citizen, Dependent income $60,000 or more	210	91.7	1.09
Prior respondent, Black/Hispanic/Asian/Hawaiian/ Other, U.S. citizen, Individual income less than $30,000	420	83.2	1.21
Prior respondent, Black/Hispanic/Asian/Hawaiian/ Other, U.S. citizen, Individual income $30,000 or more	110	91.1	1.09
Prior respondent, Black/Hispanic/Asian/Hawaiian/ Other, not U.S. citizen	60	69.6	1.49

[1]Due to small cell sizes, private, not-for-profit and private, for-profit institutions were combined.
NOTE: Predictor variables are from base-year data (NPSAS:93) with the exceptions of citizenship (B&B:93/97), phone numbers available, times answering machine encountered, and whether in field cluster (B&B:93/03).
The weight used is B03_W1. Due to model dependencies and small sample sizes, not all levels of the variables were included in the model. The denominator consists of the weighted count of sample members, excluding those who were ineligible.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

6.1.2.2 Weight adjustment for nonrespondents who refused

Of the 9,730 who were eligible and located for the B&B:93/03 sample, 590 refused. An adjustment was performed to the weight, LOCWT, which had already been adjusted for those not located, to adjust for the 590 who refused. As in the case of the adjustment for the not located, a CHAID analysis was performed on all of the predictor variables to detect important interactions. All potentially important variables were included in the initial model. There was very little change in the UWE when highly nonsignificant variables were deleted from the model, and, as a result, all variables were retained in the model.

Table 26 presents the predictor variables used in the logistic model that adjusted the weights for those who refused, and gives the weighted nonrefusal rate for those who were located and the average weight adjustment factors resulting from these variables. The weighting adjustment factor for student j is the reciprocal of the predicted response probability, or

$$ADJ3_j = 1/p_{Nref,j.}$$

The weight adjusted for those who refused was computed as

NREFWT = LOCWT * ADJ3 for the 9,140 who did not refuse
= 0 otherwise.

Table 26. Average weight adjustment factors from the logistic model used to adjust for student refusal nonresponse: 2003

Predictor variables	Number of nonrefusals	Weighted nonrefusal rate	Average weight adjustment
Total	9,140	92.2	1.07
Age			
21 or younger	2,460	92.5	1.07
22	2,500	93.3	1.06
23	1,240	91.2	1.08
24 to 27	1,320	93.1	1.06
28 or older	1,600	90.4	1.09
Race/ethnicity			
White, non-Hispanic	7,780	92.2	1.07
Black or African American, non-Hispanic	520	91.3	1.08
Hispanic	450	93.5	1.06
Asian/Native Hawaiian/Other Pacific Islander	320	92.1	1.08
American Indian/Alaska Native	60	95.0	1.05
Gender			
Male	3,910	92.4	1.07
Female	5,230	92.1	1.07
U.S. citizenship			
Yes	9,020	92.2	1.07
No	110	92.4	1.07
Attendance status			
Full time/full year: 1 institution	4,200	92.4	1.07
Full time/full year: more than 1 institution	200	93.2	1.06
Full time/part year	1,590	92.1	1.07
Part time/full year: 1 institution	1,380	92.1	1.07
Part time/full year: more than 1 institution	140	96.2	1.03
Part time/part year	1,580	91.8	1.08
Institution control			
Public	5,960	93.1	1.06
Private, not-for-profit	3,110	90.2	1.09
Private, for-profit	70	95.5	1.05
Institution region			
New England	660	90.3	1.09
Mid East	1,500	88.5	1.11
Great Lakes	1,490	91.6	1.08
Plains	800	93.1	1.06
Southeast	2,260	94.0	1.05
Southwest	1,020	94.3	1.05
Rocky Mountains	310	90.5	1.10
Far West	1,010	93.3	1.06
Outlying areas	70	97.4	1.03
Type of institution/enrollment category			
Public			
Fewer than 2,500	140	88.2	1.10
2,500–4,999	430	92.5	1.06
5,000–9,999	910	91.0	1.09

See notes at end of table.

Table 26. Average weight adjustment factors from the logistic model used to adjust for student refusal nonresponse: 2003—Continued

Predictor variables	Number of nonrefusals	Weighted nonrefusal rate	Average weight adjustment
Type of institution/enrollment category—Continued			
Public—Continued			
10,000–19,999	1,400	92.0	1.07
20,000 or more	3,060	94.4	1.05
Private, not-for-profit			
Fewer than 1,000	220	92.4	1.07
1,000–2,499	980	92.9	1.06
2,500–4,999	720	91.0	1.07
5,000–9,999	590	90.2	1.09
10,000 or more	590	85.6	1.15
Private, for-profit			
Fewer than 1,000	40	99.0	1.00
1,000 or more	40	89.4	1.10
B&B institution stratum[1]			
Public 4-year first-professional high education	370	93.8	1.05
Public 4-year first-professional low education	2,290	94.4	1.05
Private 4-year first-professional high education	680	87.4	1.12
Private 4-year first-professional low education	260	89.3	1.10
Public 4-year doctor's high education	360	92.5	1.07
Public 4-year doctor's low education	770	94.0	1.05
Private 4-year doctor's high education	120	84.0	1.16
Private 4-year doctor's low education	130	94.6	1.04
Public 4-year master's high education	310	90.0	1.10
Public 4-year master's low education	1,640	90.9	1.09
Private 4-year master's high education	140	97.6	1.01
Private 4-year master's low education	1,040	90.9	1.09
Public 4-year bachelor's high education	110	96.5	1.02
Public 4-year bachelor's low education	120	96.6	1.05
Private 4-year bachelor's high education	100	92.5	1.08
Private 4-year bachelor's low education	720	93.5	1.05
B&B student stratum			
Other students: Combined cell	500	92.5	1.08
Business majors	730	91.3	1.09
Non business majors	7,910	92.4	1.07
Applied for aid			
Yes	5,170	93.6	1.06
No	3,750	91.1	1.08
Receipt of federal aid			
Yes	3,660	93.4	1.06
No	5,460	91.6	1.08
Receipt of Pell Grant			
Yes	2,070	93.4	1.06
No	7,040	92.0	1.07
Receipt of Stafford Loan			
Yes	2,800	93.8	1.06
No	6,320	91.7	1.08

See notes at end of table.

Table 26. Average weight adjustment factors from the logistic model used to adjust for student refusal nonresponse: 2003—Continued

Predictor variables	Number of nonrefusals	Weighted nonrefusal rate	Average weight adjustment
Receipt of state aid			
Yes	1,430	94.2	1.05
No	7,690	91.9	1.07
Receipt of institution aid			
Yes	2,390	94.3	1.06
No	6,730	91.6	1.07
Receipt of any aid			
Yes	5,000	93.6	1.06
No	4,120	90.9	1.08
Prior respondent			
Respondent to either B&B:93/94 or B&B:93/97	9,020	93.4	1.07
Nonrespondent to both surveys	120	72.8	1.40
Parents' income (for dependent students)			
Less than $10,000	190	93.7	1.06
10,000–19,999	340	94.7	1.05
20,000–29,999	520	94.1	1.06
30,000–39,999	570	92.2	1.08
40,000–49,999	670	90.9	1.08
50,000–59,999	940	91.7	1.07
60,000–69,999	690	93.5	1.06
70,000–79,999	420	91.9	1.07
80,000–99,999	480	93.6	1.06
100,000 or more	660	93.0	1.07
Student's income (for independent students)			
Less than $5,000	700	93.8	1.05
5,000–9,999	710	94.9	1.05
10,000–19,999	750	91.5	1.07
20,000–29,999	450	92.6	1.06
30,000–49,999	580	89.0	1.12
50,000 or more	300	88.7	1.12
Telephone numbers available			
0 or 1	400	85.1	1.18
2	1,260	90.4	1.09
3	2,150	91.4	1.08
4	2,140	93.3	1.06
5	2,000	94.5	1.05
6	810	93.2	1.06
7 or more	360	93.1	1.06
Number times answering machine encountered			
0	2,920	97.7	1.02
1	870	93.4	1.06
More than 1	5,350	89.4	1.10
In field cluster			
Yes	4,100	91.2	1.08
No	5,040	93.2	1.06

See notes at end of table.

Table 26. Average weight adjustment factors from the logistic model used to adjust for student refusal nonresponse: 2003—Continued

Predictor variables	Number of nonrefusals	Weighted nonrefusal rate	Average weight adjustment
Interaction segments			
Prior respondent, answering machine 0 times, age 23 or less	1,890	99.3	1.01
Prior respondent, answering machine 0 times, age 24 and higher	980	96.5	1.03
Prior respondent, answering machine 1 time, in field cluster	340	90.4	1.10
Prior respondent, answering machine 1 time, not in field cluster	520	97.0	1.03
Prior respondent, answering machine more than 1 time, 0 or 1 telephone number	130	70.4	1.42
Prior respondent, answering machine more than 1 time, 2 telephone numbers	530	86.9	1.15
Prior respondent, answering machine more than 1 time, 3 or more telephone numbers	4,620	92.1	1.08
Not a prior respondent	120	72.8	1.40

[1]Due to small cell sizes, private, not-for-profit and private, for-profit institutions were combined.
NOTE: Predictor variables are from base-year data (NPSAS:93) with the exceptions of citizenship (B&B:93/97), phone numbers available, times answering machine encountered, and whether in field cluster (B&B:93/03). Due to model dependencies and small sample sizes, not all levels of the variables were included in the model. The weight used is LOCWT. The denominator used for the rate is the weighted count of eligible, located sample members.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

6.1.2.3 Weight adjustments for located nonrespondents who were not refusals

Of the 9,140 who were eligible, located, and did not refuse, 8,970 completed the B&B:93/03 interview. The remaining 170 did not respond for a reason other than a refusal (e.g., sample members who were located but who screened out calls from interviewers and, therefore, could not be reached). An adjustment was made to NREFWT to adjust for these remaining 170 students who did not respond for reasons other than refusal. As in the case of the other adjustments, a CHAID analysis was performed on all of the predictor variables to detect important interactions. All potentially important variables were included in the model. There was very little change in the unequal weighting effect when highly nonsignificant variables were deleted from the model, and as a result, all variables were retained in the model.

Table 27 presents the final predictor variables used in the logistic model that adjusted the weights for those who were interviewed and gives the weighted interview rate for those who were located and did not refuse and the average weight adjustment factors resulting from these variables. The weighting adjustment factor for student, *j*, was the reciprocal of the predicted response probability, or

$$ADJ4_j = 1/p_{R,j}$$

and the weight was computed as

BNBWT3U = NREFWT * ADJ4 for the 8,970 who responded, and

= 0 otherwise.

This final weight was rounded to the nearest integer and is denoted by BNBWT3. This weight is to be used for analyzing data collected during B&B:93/03.

Table 27. Average weight adjustment factors from the logistic model used to adjust for nonresponse other than refusal: 2003

Predictor variables	Number interviewed in B&B:93/03	Weighted interview rate	Average weight adjustment
Total	8,970	97.4	1.02
Age			
21 or younger	2,430	98.2	1.02
22	2,450	97.4	1.02
23	1,230	98.3	1.01
24 to 27	1,290	95.1	1.04
28 or older	1,560	97.5	1.02
Race/ethnicity			
White, non-Hispanic	7,660	97.7	1.02
Black or African American, non-Hispanic	500	95.5	1.04
Hispanic	440	97.7	1.02
Asian/Native Hawaiian/Other Pacific Islander	300	93.5	1.06
American Indian/Alaska Native	60	96.2	1.04
Gender			
Male	3,820	96.8	1.03
Female	5,150	97.8	1.02
U.S. citizenship			
Yes	8,870	97.8	1.02
No	100	84.4	1.26
Attendance status			
Full time/full year: 1 institution	4,140	98.4	1.01
Full time/full year: more than 1 institution	190	94.0	1.05
Full time/part year	1,560	96.0	1.04
Part time/full year: 1 institution	1,350	97.1	1.03
Part time/full year: more than 1 institution	130	93.3	1.04
Part time/part year	1,550	96.9	1.02
Institution control			
Public	5,850	97.6	1.02
Private, not-for-profit	3,050	97.7	1.02
Private, for-profit	70	80.5	1.23
Institution region			
New England	650	98.4	1.02
Mid East	1,470	96.5	1.03
Great Lakes	1,470	98.1	1.01
Plains	790	97.2	1.02
Southeast	2,210	97.4	1.03
Southwest	1,000	98.6	1.01
Rocky Mountains	310	98.7	1.01
Far West	990	95.8	1.03
Outlying areas	70	93.9	1.04
Type of institution/enrollment category			
Public			
Fewer than 2,500	140	94.9	1.04
2,500–4,999	410	95.4	1.04
5,000–9,999	900	97.4	1.02

See notes at end of table.

Table 27. Average weight adjustment factors from the logistic model used to adjust for nonresponse other than refusal: 2003—Continued

Predictor variables	Number interviewed in B&B:93/03	Weighted interview rate	Average weight adjustment
Type of institution/enrollment category—Continued			
Public—continued			
10,000–19,999	1,380	98.4	1.01
20,000 or more	3,010	97.6	1.02
Private, not-for-profit			
Fewer than 1,000	220	95.7	1.05
1,000–2,499	970	97.9	1.02
2,500–4,999	710	99.2	1.01
5,000–9,999	580	96.7	1.03
10,000 or more	580	97.7	1.02
Private, for-profit			
Fewer than 1,000	30	72.3	1.41
1,000 or more	30	94.7	1.06
B&B institution stratum[1]			
Public 4-year first-professional high education	360	98.2	1.01
Public 4-year first-professional low education	2,250	97.9	1.02
Private 4-year first-professional high education	660	97.7	1.02
Private 4-year first-professional low education	250	96.9	1.02
Public 4-year doctor's high education	350	98.9	1.01
Public 4-year doctor's low education	760	98.4	1.01
Private 4-year doctor's high education	110	95.0	1.06
Private 4-year doctor's low education	130	98.7	1.00
Public 4-year master's high education	300	97.8	1.02
Public 4-year master's low education	1,600	97.5	1.02
Private 4-year master's high education	140	99.2	1.00
Private 4-year master's low education	1,020	98.1	1.01
Public 4-year bachelor's high education	110	96.2	1.03
Public 4-year bachelor's low education	120	87.0	1.12
Private 4-year bachelor's high education	100	96.7	1.04
Private 4-year bachelor's low education	700	94.1	1.04
B&B student stratum			
Other students: Combined cell	480	95.2	1.05
Business majors	720	98.8	1.01
Nonbusiness majors	7,770	97.6	1.02
Applied for aid			
Yes	5,080	97.4	1.02
No	3,680	97.6	1.02
Receipt of federal aid			
Yes	3,590	96.7	1.03
No	5,360	97.7	1.02
Receipt of Pell Grant			
Yes	2,030	95.7	1.04
No	6,930	97.8	1.02
Receipt of Stafford Loan			
Yes	2,750	96.7	1.03
No	6,200	97.6	1.02

See notes at end of table.

Table 27. Average weight adjustment factors from the logistic model used to adjust for nonresponse other than refusal: 2003—Continued

Predictor variables	Number interviewed in B&B:93/03	Weighted interview rate	Average weight adjustment
Receipt of state aid			
Yes	1,400	97.2	1.02
No	7,550	97.4	1.02
Receipt of institution aid			
Yes	2,350	97.1	1.02
No	6,600	97.5	1.02
Receipt of any aid			
Yes	4,910	97.3	1.02
No	4,050	97.5	1.02
Prior respondent			
Respondent to either B&B:93/94 or B&B:93/97	8,860	97.5	1.02
Nonrespondent to both surveys	110	95.6	1.04
Parents' income (for dependent students)			
Less than $10,000	190	95.8	1.03
10,000–19,999	330	98.3	1.01
20,000–29,999	510	97.0	1.03
30,000–39,999	560	97.7	1.02
40,000–49,999	670	99.5	1.00
50,000–59,999	920	97.9	1.02
60,000–69,999	680	98.5	1.01
70,000–79,999	410	97.2	1.02
80,000–99,999	480	98.6	1.00
100,000 or more	650	97.6	1.02
Student's income (for independent students)			
Less than $5,000	680	93.5	1.06
5,000–9,999	690	96.5	1.03
10,000–19,999	720	94.0	1.05
20,000–29,999	440	97.7	1.01
30,000–49,999	570	99.5	1.00
50,000 or more	300	99.7	1.00
Telephone numbers available			
0 or 1	400	99.2	1.01
2	1,230	95.1	1.04
3	2,110	97.2	1.03
4	2,110	98.2	1.01
5	1,980	98.5	1.01
6	800	98.7	1.01
7	360	92.5	1.07
Number times answering machine encountered			
0	2,880	98.2	1.01
1	860	98.5	1.01
More than 1	5,230	96.8	1.03
In field cluster			
Yes	4,000	96.6	1.03
No	4,970	98.1	1.02

See notes at end of table.

Table 27. Average weight adjustment factors from the logistic model used to adjust for nonresponse other than refusal: 2003—Continued

Predictor variables	Number interviewed in B&B:93/03	Weighted interview rate	Average weight adjustment
Interaction segment			
U.S. citizen, White, 0-1 times answering machine	3,220	98.8	1.01
U.S. citizen, White, 2 or more times answering machine	4,390	97.7	1.02
U.S. citizen, race other than white	1,240	95.8	1.04
Not a U.S. citizen	100	84.4	1.26

[1]Due to small cell sizes, private, not-for-profit and private, for-profit institutions were combined.
NOTE: Predictor variables are from base-year data (NPSAS:93) with the exceptions of citizenship (B&B:93/97), and phone numbers available, times answering machine encountered, and whether in field cluster (B&B:93/03). Due to model dependencies and small sample sizes, not all levels of the variables were included in the model. The weight used is NREFWT. The denominator of the rate is the weighted count of the eligible persons who were located and did not refuse.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

6.1.3 Panel Weight

A panel (or longitudinal) weight, BNBPANL3, was also constructed for analyzing the 8,090 students who responded to all four studies:B&B:93/03, B&B93/97, B&B:93/94, and NPSAS:93. This weight was constructed by applying an additional nonresponse adjustment to the final B&B:93/03 cross-sectional weight (i.e., BNBWT3).

As for the other models, CHAID was to be used to determine the interaction segments, and the GEM modeling procedure was used to determine the adjustment factor. Table 28 presents the final predictor variables used in the logistic model that adjusted the weights for those who were not also interviewed in all three of NPSAS:93, B&B:93/94, and B&B:93/97, and gives the weighted interview rate for those who were interviewed in B&B:93/03 and the average weight adjustment factors resulting from these variables. The final weight was rounded to integer values.

Table 28. Average weight adjustment factors from the logistic model used to adjust for nonresponse to NPSAS:93, B&B:93/94 or B&B:93/97, among the respondents to B&B:93/03

Predictor variables	Number of respondents to B&B:93/03, B&B:93/94, B&B:93/97, and NPSAS:93	Weighted interview rate	Average weight adjustment
Total	8,090	84.5	1.18
Age			
21 or younger	2,190	85.5	1.16
22	2,200	83.6	1.20
23	1,100	82.8	1.21
24 to 27	1,170	84.1	1.18
28 or older	1,440	86.5	1.15
Race/ethnicity			
White, non-Hispanic	6,910	85.1	1.17
Black or African American, non-Hispanic	460	86.2	1.16
Hispanic	390	77.1	1.31
Asian/Native Hawaiian/Other Pacific Islander	280	81.4	1.24
American Indian/Alaska Native	50	91.8	1.10
Gender			
Male	3,470	85.0	1.18
Female	4,630	84.2	1.18
U.S. citizenship			
Yes	8,010	85.3	1.18
No	80	59.8	1.25
Attendance status			
Full time/full year: 1 institution	3,740	84.8	1.18
Full time/full year: more than 1 institution	190	95.5	1.05
Full time/part year	1,420	84.5	1.19
Part time/full year: 1 institution	1,220	84.1	1.19
Part time/full year: more than 1 institution	120	90.2	1.12
Part time/part year	1,380	84.8	1.18
Institution control			
Public	5,300	84.8	1.18
Private, not-for-profit	2,740	84.8	1.17
Private, for-profit	60	69.5	1.39
Institution region			
New England	560	79.2	1.26
Mid East	1,290	81.2	1.22
Great Lakes	1,310	83.2	1.20
Plains	730	88.2	1.14
Southeast	2,040	87.5	1.15
Southwest	920	84.2	1.19
Rocky Mountains	280	89.4	1.12
Far West	900	86.5	1.15
Outlying areas	60	67.5	1.42
Type of institution/enrollment category			
Public			
Fewer than 2,500	130	83.6	1.20
2,500–4,999	380	88.4	1.14

See notes at end of table.

Table 28. Average weight adjustment factors from the logistic model used to adjust for nonresponse to NPSAS:93, B&B:93/94 or B&B:93/97, among the respondents to B&B:93/03—Continued

Predictor variables	Number of respondents to B&B:93/03, B&B:93/94, B&B:93/97, and NPSAS:93	Weighted interview rate	Average weight adjustment
Type of institution/enrollment category—Continued			
Public—Continued			
5,000–9,999	830	87.5	1.14
10,000–19,999	1,260	84.8	1.18
20,000 or more	2,710	83.9	1.20
Private not-for-profit			
Fewer than 1,000	190	77.4	1.27
1,000–2,499	870	86.1	1.16
2,500–4,999	640	83.9	1.18
5,000–9,999	520	84.3	1.18
10,000 or more	520	86.8	1.15
Private for-profit			
Fewer than 1,000	30	68.4	1.50
1,000 or more	30	76.5	1.27
B&B institution stratum[1]			
Public 4-year first-professional high education	330	84.6	1.18
Public 4-year first-professional low education	2,030	84.4	1.18
Private 4-year first-professional high education	590	84.0	1.18
Private 4-year first-professional low education	230	86.9	1.15
Public 4-year doctor's high education	320	87.9	1.14
Public 4-year doctor's low education	690	84.0	1.19
Private 4-year doctor's high education	100	84.3	1.19
Private 4-year doctor's low education	110	77.5	1.28
Public 4-year master's high education	270	84.7	1.17
Public 4-year master's low education	1,450	84.0	1.18
Private 4-year master's high education	130	85.1	1.12
Private 4-year master's low education	920	85.7	1.16
Public 4-year bachelor's high education	100	83.3	1.22
Public 4-year bachelor's low education	110	94.2	1.06
Private 4-year bachelor's high education	90	79.4	1.24
Private 4-year bachelor's low education	630	82.2	1.20
B&B student stratum			
Other students: Combined cell	430	87.0	1.15
Business majors	640	85.2	1.17
Nonbusiness majors	7,020	83.7	1.18
Applied for aid			
Yes	4,660	87.0	1.15
No	3,260	82.3	1.22
Receipt of federal aid			
Yes	3,310	87.6	1.14
No	4,780	83.2	1.20
Receipt of Pell Grant			
Yes	1,880	88.0	1.14
No	6,220	84.0	1.19

See notes at end of table.

Table 28. Average weight adjustment factors from the logistic model used to adjust for nonresponse to NPSAS:93, B&B:93/94 or B&B:93/97, among the respondents to B&B:93/03—Continued

Predictor variables	Number of respondents to B&B:93/03, B&B:93/94, B&B:93/97, and NPSAS:93	Weighted interview rate	Average weight adjustment
Receipt of Stafford Loan			
Yes	2,540	87.9	1.14
No	5,550	83.6	1.20
Receipt of state aid			
Yes	1,290	87.4	1.14
No	6,810	84.3	1.19
Receipt of institution aid			
Yes	2,180	88.1	1.14
No	5,920	83.7	1.19
Receipt of any aid			
Yes	4,500	87.3	1.15
No	3,590	82.2	1.22
Parent's income (for dependent students)			
Less than $10,000	170	90.0	1.11
10,000–19,999	300	86.2	1.19
20,000–29,999	480	89.6	1.14
30,000–39,999	510	87.4	1.16
40,000–49,999	610	83.6	1.21
50,000–59,999	830	81.0	1.25
60,000–69,999	620	84.6	1.19
70,000–79,999	370	81.9	1.24
80,000–99,999	430	85.4	1.22
100,000 or more	600	89.9	1.14
Student's income (for independent students)			
Less than $5,000	630	85.7	1.19
5,000–9,999	640	90.6	1.11
10,000–19,999	660	84.1	1.17
20,000–29,999	410	85.7	1.17
30,000–49,999	530	85.3	1.17
50,000 or more	260	89.5	1.12
Telephone numbers available			
0 or 1	340	75.8	1.33
2	1,070	78.3	1.29
3	1,880	84.3	1.18
4	1,930	87.3	1.14
5	1,810	86.7	1.15
6	740	87.8	1.14
7	330	87.2	1.14
Number times answering machine encountered			
0	2,600	85.9	1.16
1	780	85.1	1.16
More than 1	4,720	83.8	1.19
In field cluster			
Yes	3,560	82.2	1.22
No	4,540	86.7	1.15

See notes at end of table.

Table 28. Average weight adjustment factors from the logistic model used to adjust for nonresponse to NPSAS:93, B&B:93/94 or B&B:93/97, among the respondents to B&B:93/03—Continued

Predictor variables	Number of respondents to B&B:93/03, B&B:93/94, B&B:93/97, and NPSAS:93	Weighted interview rate	Average weight adjustment
Interaction segment			
0–2 phone numbers, Pell grant recipient	300	82.6	1.23
0–2 phone numbers, not a Pell grant recipient	1,100	76.9	1.32
3 phone numbers, in field cluster, 22 years old or younger	480	84.3	1.18
3 phone numbers, in field cluster, age 23–27	200	74.5	1.35
3 phone numbers, in field cluster, age 28 or older	130	83.3	1.20
3 phone numbers, not in field cluster, received institution aid	310	93.3	1.08
3 phone numbers, not in field cluster, no institution aid	760	85.2	1.18
4 or more phone numbers, received Stafford loan	1,570	91.8	1.09
4 or more phone numbers, no Stafford loan, received institution aid	620	90.1	1.12
4 or more phone numbers, no Stafford loan, no inst aid	2,620	85.0	1.18

[1]Due to small cell sizes, private, not-for-profit and private, for-profit institutions were combined.
NOTE: Predictor variables are from base-year data (NPSAS:93) with the exceptions of citizenship (B&B:93/97), phone numbers available, times answering machine encountered, and whether in field cluster (B&B:93/03). Due to model dependencies and small sample sizes, not all levels of the variables were included in the model.
The weight used is BNBWT3. The denominator is the weighted count of eligible respondents to B&B:93/03.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

6.1.4 Weight Distributions and Unequal Weighting Effects

The distributions of the weight adjustment factors for the B&B:93/03 analysis weights and the distributions of the initial, intermediate, and final weights along with their unequal weighting design effects are presented in Tables 29 and 30.

Table 29. Distribution of values for the B&B:93/03 weight adjustment factors

Quantile	Location ADJ2	Refusal ADJ3	Nonresponse ADJ4	Longitudinal ADJ5
Minimum percent	1.00	1.00	1.00	1.00
1	1.00	1.00	1.00	1.02
5	1.01	1.00	1.00	1.04
10	1.01	1.00	1.00	1.05
25	1.03	1.01	1.00	1.09
Median percent	1.05	1.04	1.00	1.15
75	1.09	1.09	1.01	1.24
90	1.15	1.16	1.04	1.35
95	1.22	1.23	1.10	1.43
99	1.43	1.51	1.37	1.57
Maximum	1.98	1.89	2.00	1.93

SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

Table 30. Distribution of initial, intermediate, and final weights for B&B:93/03

	Initial weight	Intermediate weights		Cross-sectional analysis weight	Longitudinal analysis weight
Quantile	BB03_W1	Location LOCWT	Refusal NREFWT	BNBWT3	BNBPANL3
Minimum percent	2	2	2	2	3
1	22	24	25	25	28
5	38	41	42	43	47
10	49	53	56	57	63
25	69	73	76	77	87
Median percent	86	91	95	96	112
75	107	112	121	123	146
90	229	243	257	259	277
95	334	367	385	395	395
99	454	496	573	605	623
Maximum percent	2,446	2,653	2,654	2,861	2,131
Design effect	1.713	1.767	1.839	1.846	1.679

SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

6.2 Variance Estimation

For probability-based sample surveys, most estimates are nonlinear statistics. For example, a mean or proportion is calculated as *Ewy/Ew*, which is nonlinear because the denominator is a survey estimate of the unknown population total. In this situation, the variances of the estimates cannot be expressed in closed form. Two common procedures for estimating the variances of nonlinear survey statistics are Taylor series linearization procedures and replication methods. The replication method used in B&B:93/03 is balanced repeated replication (BRR). BRR is used because of its superiority for the estimation of the variances of quantiles, such as medians. The subsections below discuss the Taylor series and BRR methods of variance estimation for B&B:93/03.

6.2.1 Taylor Series

The Taylor series variance estimation procedure is a well-known technique for estimating variances of nonlinear statistics. The procedure substitutes the first-order Taylor series approximation of the nonlinear statistic into the variance formula based on the sampling design. Woodruff (1971) presents the mathematical formulation of this procedure. For stratified, multistage sampling designs, the Taylor series procedure requires analysis strata and analysis replicates based on the first stage sampling design. Since the B&B:93/03 sample is a subsample of the B&B:93/97 eligibles, the B&B:93/03 Taylor series strata and PSUs were derived from the Taylor series strata and PSUs for the B&B:93/97 sample.

The B&B:93/97 variance estimation strata and PSUs were defined as follows. Using the NPSAS:93 Taylor series strata and PSUs (where available), Taylor series strata (TAYSTRBB) and PSUs (TAYREPBB) were defined for the 11,160 B&B:93/97 eligible sample students.

Missing values were resolved using prior knowledge of how the variables were assigned for NPSAS:93 and other variables on the file. Once each sample student had non-missing Taylor series strata and PSUs, strata and PSUs were collapsed in order to get at least four respondents per PSU. This process resulted in variance estimation strata and PSUs for analyzing the B&B:93/97 data; these variables are named TAYSTR97 and TAYREP97.

To define variance estimation strata and PSUs for B&B:93/03, these B&B:93/97 Taylor series and PSUs were collapsed in order to get at least four respondents per PSU. The same collapsing rules that were used to obtain the Taylor series strata and PSUs for the B&B:93/97 respondents were used.

The following summarizes the variable names for the weights, analysis strata, and analysis replicates on the B&B:93/03 data file for use with Taylor series variance estimation.

TAYSTR03 B&B:93/03 analysis strata

TAYREP03 B&B:93/03 analysis replicate

BNBWT3 B&B:93/03 analysis weight for 2003 respondents, for cross-sectional analyses

BNBPANL3 B&B:93/03 panel weight, for respondents to all four studies: NPSAS:93, B&B:93/94, B&B:93/97, and B&B:93/03

6.2.2 Balanced Repeated Replication

BRR is one of two replication techniques commonly used to estimate the variances of survey statistics computed from complex sample surveys. (The other commonly used replication technique is the jackknife replication technique.) Wölter (1985) reviews both the Taylor series and replication techniques.

The BRR method is designed for a survey with L primary sampling strata and two PSUs selected per stratum. A half-sample replicate is formed by selecting one PSU from each stratum. For any given sample, there are 2^L such half-samples. If $\bar{y}_{st,\alpha}$ represents the estimate of the population mean calculated from the α-th replicate and $\bar{y}_{st}$ represents the stratified mean from the full sample, then the mean of $(\bar{y}_{st,\alpha} - \bar{y}_{st})^2$ over all 2^L half samples is identical to the textbook stratified variance estimator. BRR is essentially a method for selecting a set of k "balanced" replicates where k is much smaller than 2^L so that this same property holds for the set of k replicates (see chapter 3 of Wölter 1985). The BRR variance estimate is then computed as

$$\mathrm{Var}_{\mathrm{BRR}}\ (\bar{y}_{st}) = \sum_{\alpha=1}^{k} \frac{(\bar{y}_{st,\alpha} - \bar{y}_{st})^2}{k} .$$

BRR weights were computed for B&B:93/03 because of concern that the variances for medians and other quartiles may not be appropriate when computed using either Taylor series or jackknife methods. The Taylor series approach estimates the cumulative distribution function at

several points and then estimates variances for quartiles through inverse interpolation (see Francisco and Fuller 1991). Because these results depend on the points at which the cumulative distribution function and its variances are evaluated, they are subjective and require considerable care by the user. Likewise, jackknife methods are inconsistent for estimating the variances of nonsmooth functions, such as quartiles (see chapter 3 of Efron 1982); as the sample size increases, the estimates do not converge to the true value. Moreover, the resulting jackknife variance estimator has only two degrees of freedom, irrespective of the sample size.

As mentioned above, the BRR method is designed for surveys with two PSUs per stratum. Because the NPSAS:93 was not a two-PSU-per-stratum design, the first task was to approximate the design for variance estimation purposes as one with two analysis PSUs per stratum. Fortunately, that problem was solved when the NPSAS:93 jackknife weights were computed. As explained in section 7.4.2 of the NPSAS:93 Methodology Report, when computing the jackknife weights, a set of 42 pseudo-strata were developed. Instead of continuing with jackknife weights, BRR weights were computed for B&B:93/03 because of the superiority of BRR variance estimation for medians and other quantiles.

The $L = 42$ pseudo-strata defined for NPSAS:93 were used to compute BRR weights based on the initial weights for the B&B:93/03 follow-up, namely BB03_W1. Wölter (1985) explains that to achieve "full orthogonal balance," k half-sample replicates should be used where $k > L$ and k is a multiple of 4. Since 11*4 = 44, $k = 44$ will be used. As Wölter further explains, any 44x44 Hadamard matrix can be used to define the 44 balanced half-samples. In particular, the 44 rows can be used to represent the 44 BRR replicates and any 42 columns can be used to represent the 42 NPSAS:93 pseudo-strata. A 44x44 Hadamard matrix was used to compute the BRR weights for B&B:93/03. Two columns were randomly selected to be deleted in order to identify 44 BRR replicate samples, as discussed below.

Using Wölter's notation (with rows and columns reversed), let $\delta_h^{(\alpha)}$ denote the element of the 44x44 Hadamard matrix in row h and column α. The "+1" and "-1" elements of the matrix will be used to define 44 initial balanced replicate weights from B03_01 and the NPSAS:93 jackknife replicate and stratum variables, JACKREP and JACKSTR, as follows:

+1 → the α-th BRR replicate contains the observations in the first pseudo-replicate from pseudo-stratum h (BRRWTα = 2 times B03_01 if JACKREP = 1; BRRWTα = 0 if JACKREP = 2); and

–1 → the α-th BRR replicate contains the observations in the second pseudo-replicate from pseudo-stratum h (BRRWTα = 2 times B03_01 if JACKREP = 2; BRRWTα = 0 if JACKREP = 1).

From each of the 44 BRR initial replicate weights defined in this manner, the final BRR replicate weights were computed using exactly the same weight adjustment procedures that had been implemented for the full B&B:93/03 sample.

Two sets of BRR weights were computed, one corresponding to the cross-sectional weight, and the other corresponding to the panel weight. The final BRR weights, rounded to integer values, are named as follows.

B03BRR01 – B03BRR44	BRR weights for the 2003 respondents, to be used for cross-sectional analyses; and
B3PBRR01 – B3PBRR44	BRR weights for respondents to the 1993, 1994, 1997, and 2003 surveys, to be used for longitudinal analyses.

In addition, BRR weights corresponding to the initial B&B weight, BNBWTO, were constructed for analyzing the full B&B cohort. These weights are

B00BRR01 – B00BRR44	BRR weights corresponding to BNBWTO, for the 11,190 students in the B&B cohort.

Table 31 summarizes the variables and how they are used in selected software packages that allow for BRR variance estimation (SUDAAN, WesVar, and STATA).

Table 31. Analysis weight, strata, and replicate variables that are available from B&B:93/03

	Type of analysis	
	Cross-sectional	Longitudinal
	B&B:93/03 respondents	Respondents to all four of NPSAS:93, B&B:93/94, B&B:93/97, B&B:93/03
Weight variables for estimates	BNBPWT3	BNBPANL3
	Taylor series variance estimation	
Variables denoting analysis strata and replicates	TAYSTR03 TAYREP03	TAYSTR03 TAYREP03
SUDAAN	DESIGN = WR WEIGHT BNBWT3; NEST TAYSTR03 TAYREP03;	DESIGN = WR WEIGHT BNBPANL3; NEST TAYSTR03 TAYREP03;
STATA	SVYSET [PWEIGHT = BNBWT3], STRATA (TAYSTR03), PSU (TAYREP03)	SVYSET [PWEIGHT = BNBPANL3], STRATA (TAYSTR03), PSU (TAYREP03)
SAS survey data analysis procedures	WEIGHT BNBWT3; STRATA TAYSTR03; CLUSTER TAYREP03	WEIGHT BNBPANL3; STRATA TAYSTR03; CLUSTER TAYREP03
	BRR variance estimation	
Replicate weight variables	B03BRR01 – B03BRR44	B3PBRR01 – B3PBRR44
SUDAAN	DESIGN = BRR WEIGHT BNBWT3; REPWGT B03BRR01 – B03BRR44	DESIGN = BRR WEIGHT BNBPANL3; REPWGT B3PBRR01 – B3PBRR44
WESVAR	Method BRR Full sample weight BNBWT3 Replicates B03BRR01 – B03BRR44	Method BRR Full sample weight BNBPANL3 Replicates B3PBRR01 – B3PBRR44

SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

6.3 Accuracy of Estimates

The accuracy of survey statistics is affected by both random and nonrandom errors. Random errors reduce the precision of survey statistics, while nonrandom errors result in bias (i.e., estimates that do not converge to the true population parameter as the sample size increases without limit).

The sources of error in a survey are often dichotomized as sampling and nonsampling errors. Sampling error refers to the error that occurs simply because the survey is based on a sample of population members, rather than the entire population. All other types of errors are nonsampling errors, including survey nonresponse (due to the inability to contact sampling members, their refusal to participate in the study, etc.) and measurement errors, such as the errors that occur because the intent of survey questions was not clear to the respondent, because the respondent had insufficient knowledge to answer correctly, or because the data were not captured correctly (e.g., because of recording, editing, or data entry errors).

Sampling errors are primarily random errors for well-designed surveys, like B&B:93/03. However, nonrandom errors may occur if the sampling frame does not provide complete coverage of the target population. The B&B:93/03 survey instrument and data collection procedures were subjected to thorough development and testing to minimize nonsampling errors because these errors are difficult to quantify and are likely to be nonrandom errors.

In section 6.3.1, design effect calculations are described, and sampling errors and design effects for some B&B:93/03 estimates are computed and presented for a variety of domains. Section 6.3.2 presents an analysis which compares the B&B:93/03 nonrespondents and respondents using characteristics known for both nonrespondents and respondents to look for characteristics that may be related to bias due to unit nonresponse and for variables that might be used in the weighting to adjust for this potential bias. An analysis examining the effect that the nonresponse adjustment had on this potential bias is also presented.

6.3.1 Measures of Precision: Standard Errors and Design Effects

The cumulative effect of random errors on the precision of a survey statistic is measured by the standard error of that statistic. The standard error of a statistic is the estimated standard deviation of the sampling distribution of the statistic over repeated samples of the same size using the same sampling design. Hence, the standard error of a survey statistic depends not only on the natural variability of the observations in the population and on the sample size, but also on the characteristics of the sampling design. Features of the sampling design that affect the sampling variance of a survey statistic (the square of the standard error) include stratification, multistage or cluster sampling, and unequal sampling rates. Stratification can increase precision if outcomes are more homogeneous within strata than between strata, but the other survey design features usually decrease precision. Moreover, statistical adjustment of the analysis weights to reduce the potential for bias due to nonresponse often decreases precision since these adjustments often cause an increase in the variability of the analysis weights.

The cumulative effect of the various factors affecting the precision of a survey statistic is often modeled as the survey design effect. The design effect, designated as DEFF, is defined as the ratio of the sampling variance of the statistic under the actual sampling design divided by the variance that would be expected for a simple random sample of the same size. The square root of the design effect (also called the root design effect, and designated as DEFT) is also useful. The following formulas define the design effects and root design effects:

$$\mathrm{DEFF}(\hat{\theta}) = \frac{\mathrm{Var}(\hat{\theta})}{\mathrm{Var}_{\mathrm{SRS}}(\hat{\theta})}$$

$$\mathrm{DEFT}(\hat{\theta}) = \frac{\mathrm{SE}(\hat{\theta})}{\mathrm{SE}_{\mathrm{SRS}}(\hat{\theta})}$$

In these formulas, $\hat{\theta}$ represents the survey statistic of interest. Hence, the design effect is unity (1.00), by definition, for simple random samples. For most practical sampling designs, the

survey design effect is greater than unity, reflecting that the precision is less than could be achieved with a simple random sampling of the same size (if such a design were practical). The size of the survey design effect depends largely on the sample size and intracluster correlation within the primary sampling units (e.g., number of students per institution and within-institution correlations). Hence, statistics that are based on observations that are highly correlated within institutions will have higher design effects for B&B:93/03.

The simple random sample variance used in this report is computed conditional on the sample size of the analysis domain. Specifically, if n_d is the respondent sample size in the domain and $\hat{\theta}_d$ is the weighted estimate of the proportion for the domain, then the simple random sample variance is computed as $\text{Var}_{\text{SRS}}(\hat{\theta}_d) = \frac{\hat{\theta}_d(1-\hat{\theta}_d)}{n_d}$.

To provide an approximate characterization of the precision with which B&B:93/03 survey statistics can be estimated, a series of tables has been prepared to provide estimates of key statistics, their standard errors, and the estimated survey design effects. These are presented in appendix H for a variety of survey estimates for domains defined by

- all respondents;
- race/ethnicity;
- gender, and
- control of the base year school (public or private)

Tables in this appendix give the percentage estimates, the design based standard errors (produced using BRR and SUDAAN [2001]), the denominator sample size, and DEFF and DEFT. The tables also give the mean, minimum, and maximum values of DEFF and DEFT for each domain. The tables present design effect results using both the cross-sectional and panel weights.

6.3.2 Measures of Bias

6.3.2.1 Nonresponse bias analysis

Unit nonresponse causes bias in survey estimates when the outcomes of respondents and nonrespondents are different. A bias analysis was conducted to determine if any variables are significantly biased due to nonresponse. Three types of nonresponse bias analyses were considered:

- nonrespondents versus respondents;
- early refusals who were later converted to respondents versus other respondents; and
- late respondents (those who responded August 1, 2003 or later) versus earlier respondents.

Persons who initially refused and those who responded later in the interview period, may have nonresponse-like behavior.

For the first of these, respondents and nonrespondents were characterized by comparing the weighted[13] percentage of respondents with the weighted percentage of nonrespondents for each category of important characteristics known for both respondents and nonrespondents. Characteristics used in the analysis included those used in the weight adjustments. As noted earlier, not all levels of all variables were included in the response adjustments due to model dependencies and small sample sizes. Statistical tests (*t* tests) were performed to determine whether the difference between respondents and nonrespondents is significant at the 5 percent level. Table 32 compares the demographic characteristics of respondents and nonrespondents. This table shows that the distributions of demographic characteristics such as race/ethnicity, citizenship, control of base-year institution, region of the base-year institution, whether the student applied for aid in the base year, receipt of federal, state, and/or institution aid, whether the student responded to prior rounds of data collection, and locator information (telephone numbers available, number of times an answering machine was encountered, and whether the student lives in a field cluster) are significantly different for respondents and nonrespondents.

Table 33 performs a similar analysis, but compares demographic characteristics of those respondents who initially refused but were later converted to respondents with other respondents. This comparison was made because the refusals who were converted are likely similar to the refusal nonrespondents who were not converted. This analysis shows few differences in the distributions of demographic variables for refusals who were later converted and other respondents. Distributions are significantly different, however, for the following characteristics: receipt of institution aid in the base year, whether the student responded to prior rounds of data collection, and the number of times an answering machine was encountered.

[13] The unrounded base weight, BB03_W1U, was used. Missing values were excluded for the demographic variables.

Table 32. Comparison of B&B:93/03 respondents and nonrespondents

Demographic characteristics	Respondents		Nonrespondents		Full sample	
	Sample size	Percent estimate[1]	Sample size	Percent Estimate[1]	Sample size	Percent Estimate[1]
Age						
21 or younger	2,430	25.74	320	22.19	2,750	25.16
22	2,450	27.26	360	26.08	2,810	27.06
23	1,230	13.70	220	13.84	1,440	13.72
24 to 27	1,290	14.24	220	16.62	1,510	14.63
28 or older	1,560	19.06	310	21.28	1,870	19.43
Race/ethnicity						
White, non-Hispanic	7,660	85.39	1,100	75.19*	8,760	83.71
Black or African American, non-Hispanic	500	5.30	120	9.74*	620	6.03
Hispanic	440	5.04	100	5.89	530	5.18
Asian/Native Hawaiian/Other Pacific Islander	300	3.68	100	8.70*	400	4.51
American Indian/Alaska Native	60	0.58	10	0.47	70	0.56
Gender						
Male	3,820	44.39	670	48.21	4,490	45.02
Female	5,150	55.61	760	51.79	5,910	54.98
U.S. citizenship						
Yes	8,870	98.27	1,370	93.09*	10,240	97.42
No	100	1.73	50	6.91*	150	2.58
Attendance status						
Full time/full year: 1 institution	4,140	45.62	620	41.05	4,760	44.87
Full time/full year: more than 1 institution	190	0.90	30	1.37	230	0.97
Full time/part year	1,560	17.38	270	19.47	1,830	17.72
Part time/full year: 1 institution	1,350	16.64	220	17.02	1,560	16.70
Part time/full year: more than 1 institution	130	0.86	20	0.85	150	0.86
Part time/part year	1,550	18.61	250	20.25	1,800	18.88
Institution control						
Public	5,850	67.64	910	62.33*	6,760	66.76
Private, not-for-profit	3,050	31.12	500	34.43	3,550	31.67
Private, for-profit	70	1.23	20	3.25	90	1.57
Institution region						
New England	650	7.02	100	7.12	750	7.04
Mid East	1,470	15.79	280	21.72*	1,740	16.77
Great Lakes	1,470	17.70	230	15.97	1,690	17.41
Plains	790	9.40	80	6.61	870	8.94
Southeast	2,210	23.46	380	23.36	2,590	23.44
Southwest	1,000	11.59	150	9.02	1,150	11.17
Rocky Mountains	310	3.16	40	3.19	350	3.17
Far West	990	11.28	160	12.47	1,150	11.48
Outlying areas	70	0.59	20	0.54	90	0.58
Type of institution/enrollment category						
Public						
Fewer than 2,500	140	0.87	30	1.26	170	0.94
2,500–4,999	410	3.63	70	4.10	490	3.70
5,000–9,999	900	9.76	150	9.48	1,050	9.71
10,000–19,999	1,380	17.17	210	16.26	1,590	17.02
20,000 or more	3,010	36.28	450	31.29	3,460	35.46

See notes at end of table.

Table 32. Comparison of B&B:93/03 respondents and nonrespondents—Continued

Demographic characteristics	Respondents Sample size	Respondents Percent estimate[1]	Nonrespondents Sample size	Nonrespondents Percent estimate[1]	Full sample Sample size	Full sample Percent estimate[1]
Type of institution/enrollment category—Continued						
Private, not-for-profit						
Fewer than 1,000	220	1.72	30	2.31	250	1.82
1,000–2,499	970	8.78	150	7.90	1,110	8.64
2,500–4,999	710	7.07	100	6.69	810	7.01
5,000–9,999	580	6.76	100	7.50	670	6.88
10,000 or more	580	6.72	120	9.96	690	7.26
Private, for-profit						
Fewer than 1,000	30	0.76	10	2.19	40	1.00
1,000 or more	30	0.48	10	1.07	50	0.58
B&B institution stratum[2]						
Public 4-year first-professional high education	360	4.26	50	3.17	410	4.08
Public 4-year first-professional low education	2,250	25.97	320	21.51	2,570	25.23
Private 4-year first-professional high education	660	8.35	140	11.75	800	8.91
Private 4-year first-professional low education	250	2.32	40	3.25	300	2.48
Public 4-year doctor's high education	350	4.72	50	3.90	410	4.58
Public 4-year doctor's low education	760	9.12	120	7.62	880	8.87
Private 4-year doctor's high education	110	1.13	30	2.15	140	1.30
Private 4-year doctor's low education	130	1.09	20	0.74	150	1.03
Public 4-year master's high education	300	3.51	50	3.45	360	3.50
Public 4-year master's low education	1,600	17.77	280	20.34	1,880	18.19
Private 4-year master's high education	140	0.67	20	0.47	160	0.64
Private 4-year master's low education	1,020	10.61	160	10.25	1,170	10.55
Public 4-year bachelor's high education	110	0.53	10	0.29	120	0.49
Public 4-year bachelor's low education	120	1.77	20	2.05	140	1.81
Private 4-year bachelor's high education	100	0.80	10	0.65	120	0.78
Private 4-year bachelor's low education	700	7.38	110	8.40	810	7.55
B&B student stratum						
Other students: Combined cell	480	17.23	90	18.94	570	17.51
Business majors	720	16.40	140	16.18	860	16.37
Nonbusiness majors	7,770	66.37	1,210	64.88	8,970	66.12
Applied for aid						
Yes	5,080	52.36	770	47.71*	5,850	51.61
No	3,680	47.64	610	52.29*	4,290	48.39
Receipt of federal aid						
Yes	3,590	32.97	570	32.59	4,160	32.91
No	5,360	67.03	860	67.41	6,220	67.09
Receipt of Pell Grant						
Yes	2,030	17.17	340	18.76	2,370	17.43
No	6,930	82.83	1,080	81.24	8,010	82.57
Receipt of Stafford Loan						
Yes	2,750	25.30	430	24.64	3,180	25.19
No	6,200	74.70	1,000	75.36	7,200	74.81
Receipt of state aid						
Yes	1,400	12.62	210	10.75	1,610	12.31
No	7,550	87.38	1,220	89.25	8,770	87.69
Receipt of institution aid						
Yes	2,350	22.16	300	17.80*	2,660	21.44
No	6,600	77.84	1,120	82.20*	7,720	78.56

See notes at end of table.

Table 32. Comparison of B&B:93/03 respondents and nonrespondents—Continued

Demographic characteristics	Respondents Sample size	Respondents Percent estimate[1]	Nonrespondents Sample size	Nonrespondents Percent estimate[1]	Full sample Sample size	Full sample Percent estimate[1]
Receipt of any aid						
Yes	4,910	48.95	740	43.97*	5,650	48.13
No	4,050	51.05	690	56.03*	4,730	51.87
Prior respondent						
Respondent to either B&B:93/94 or B&B:93/97	8,860	96.30	1,320	84.37*	10,180	94.32
Nonrespondent to both surveys	110	3.70	110	15.63*	220	5.68
Parents' income (for dependent students)						
Less than $10,000	190	1.91	40	1.89	220	1.91
10,000–19,999	330	3.09	50	2.77	390	3.03
20,000–29,999	510	5.11	80	4.61	590	5.03
30,000–39,999	560	6.00	80	5.75	640	5.96
40,000–49,999	670	7.38	80	6.65	750	7.26
50,000–59,999	920	11.46	170	12.30	1,090	11.60
60,000–69,999	680	8.16	100	7.00	780	7.97
70,000–79,999	410	4.80	60	4.50	470	4.75
80,000–99,999	480	5.45	60	4.76	540	5.33
100,000 or more	650	7.51	80	6.40	730	7.33
Student's income (for independent students)						
Less than $5,000	680	6.88	130	9.00	820	7.22
5,000–9,999	690	6.77	100	6.61	800	6.74
10,000–19,999	720	7.92	150	10.72	870	8.38
20,000–29,999	440	5.36	80	5.59	520	5.39
30,000–49,999	570	7.85	90	7.16	660	7.74
50,000 or more	300	4.36	50	4.30	340	4.35
Telephone numbers available						
0 or 1	400	4.80	90	8.36*	490	5.39
2	1,230	14.42	230	18.58	1,460	15.11
3	2,110	23.37	320	22.70	2,430	23.26
4	2,110	22.97	300	19.33	2,400	22.37
5	1,980	21.58	250	16.24*	2,230	20.70
6	800	8.77	140	8.44	940	8.71
7 or more	360	4.09	90	6.33	450	4.46
Number times answering machine encountered						
0	2,880	32.25	220	16.98*	3,110	29.72
1	860	9.24	90	6.14*	950	8.73
More than 1	5,230	58.51	1,110	76.88*	6,340	61.55
In field cluster						
Yes	4,000	46.74	700	53.01*	4,700	47.78
No	4,970	53.26	730	46.99*	5,700	52.22

*Difference between respondents and nonrespondents is significant at the $p < 0.05/(c-1)$ level, where c is the number of categories within the primary variable.
[1]The weight used is BB03_W1U (unrounded initial weight for B&B:03).
[2]Due to small cell sizes, private, not-for-profit and private, for-profit institutions were combined.
NOTE: Characteristics are from base-year data (NPSAS:93) with the exceptions of citizenship (B&B:93/97), and phone numbers available, times answering machine encountered, and whether in field cluster (B&B:93/03). Missing values are excluded from the table. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

Table 33. Comparison of B&B:93/03 converted refusals and other respondents

Demographic characteristics	Converted refusals		Other respondents		Total respondents	
	Sample size	Percent estimate[1]	Sample size	Percent estimate[1]	Sample size	Percent estimate[1]
Age						
21 or younger	120	21.84	2,300	26.00	2,430	25.74
22	110	25.37	2,330	27.38	2,450	27.26
23	80	16.84	1,150	13.50	1,230	13.70
24 to 27	80	13.16	1,210	14.31	1,290	14.24
28 or older	90	22.79	1,470	18.82	1,560	19.06
Race/ethnicity						
White, non-Hispanic	430	89.18	7,240	85.14	7,660	85.39
Black or African American, non-Hispanic	20	5.15	480	5.31	500	5.30
Hispanic	20	2.93	420	5.18	440	5.04
Asian/Native Hawaiian/Other Pacific Islander	10	1.98	290	3.80	300	3.68
American Indian/Alaska Native	10	0.75	50	0.57	60	0.58
Gender						
Male	220	47.47	3,600	44.19	3,820	44.39
Female	270	52.53	4,880	55.81	5,150	55.61
U.S. citizenship						
Yes	480	98.41	8,390	98.26	8,870	98.27
No	10	1.59	90	1.74	100	1.73
Attendance status						
Full time/full year: 1 institution	220	42.98	3,920	45.79	4,140	45.62
Full time/full year: more than 1 institution	10	1.14	180	0.88	190	0.90
Full time/part year	80	16.57	1,480	17.43	1,560	17.38
Part time/full year: 1 institution	80	18.44	1,260	16.52	1,350	16.64
Part time/full year: more than one institution	10	0.99	120	0.86	130	0.86
Part time/part year	80	19.88	1,470	18.52	1,550	18.61
Institution control						
Public	320	65.68	5,530	67.77	5,850	67.64
Private, not-for-profit	160	32.72	2,890	31.02	3,050	31.12
Private, for-profit	10	1.60	60	1.21	70	1.23
Institution region						
New England	40	7.95	610	6.96	650	7.02
Mid East	90	18.67	1,380	15.60	1,470	15.79
Great Lakes	90	19.62	1,380	17.57	1,470	17.70
Plains	50	8.94	750	9.43	790	9.40
Southeast	110	20.97	2,100	23.62	2,210	23.46
Southwest	60	13.32	940	11.48	1,000	11.59
Rocky Mountains	10	2.20	290	3.23	310	3.16
Far West	50	8.21	940	11.49	990	11.28
Outlying areas	#	0.11	70	0.62	70	0.59
Type of institution/enrollment category						
Public						
Fewer than 2,500	#	0.30	130	0.91	140	0.87
2,500–4,999	30	4.02	380	3.60	410	3.63
5,000–9,999	50	10.16	840	9.73	900	9.76
10,000–19,999	80	17.58	1,300	17.14	1,380	17.17
20,000 or more	150	33.70	2,860	36.45	3,010	36.28

See notes at end of table.

Table 33. Comparison of B&B:93/03 converted refusals and other respondents—Continued

Demographic characteristics	Converted refusals		Other respondents		Total respondents	
	Sample size	Percent estimate[1]	Sample size	Percent estimate[1]	Sample size	Percent estimate[1]
Type of institution/enrollment category						
Private, not-for-profit						
Fewer than 1,000	#	0.38	210	1.81*	220	1.72
1,000–2,499	50	9.83	920	8.71	970	8.78
2,500–4,999	50	8.21	670	7.00	710	7.07
5,000–9,999	40	10.36	540	6.52	580	6.76
10,000 or more	30	3.87	550	6.91	580	6.72
Private, for-profit						
Fewer than 1,000	#	0.56	30	0.77	30	0.76
1,000 or more	#	1.04	30	0.44	30	0.48
B&B institution stratum[2]						
Public 4-year first-professional high education	20	5.96	340	4.14	360	4.26
Public 4-year first-professional low education	110	22.58	2,140	26.20	2,250	25.97
Private 4-year first-professional high education	30	6.52	630	8.47	660	8.35
Private 4-year first-professional low education	10	1.54	240	2.38	250	2.32
Public 4-year doctor's high education	10	2.77	340	4.84	350	4.72
Public 4-year doctor's low education	50	9.42	710	9.10	760	9.12
Private 4-year doctor's high education	10	1.64	110	1.09	110	1.13
Private 4-year doctor's low education	10	1.61	120	1.06	130	1.09
Public 4-year master's high education	20	2.99	290	3.55	300	3.51
Public 4-year master's low education	100	21.28	1,500	17.54	1,600	17.77
Private 4-year master's high education	#	0.52	140	0.68	140	0.67
Private 4-year master's low education	60	11.16	960	10.57	1,020	10.61
Public 4-year bachelor's high education	10	0.62	100	0.52	110	0.53
Public 4-year bachelor's low education	#	0.09	120	1.88*	120	1.77
Private 4-year bachelor's high education	10	0.81	100	0.80	100	0.80
Private 4-year bachelor's low education	40	10.51	660	7.18	700	7.38
B&B student stratum						
Other students: Combined cell	30	17.80	450	17.19	480	17.23
Business majors	60	22.73	670	15.99*	720	16.40
Nonbusiness majors	390	59.47	7,370	66.82*	7,770	66.37
Applied for aid						
Yes	280	50.76	4,810	52.47	5,080	52.36
No	200	49.24	3,480	47.53	3,680	47.64
Receipt of federal aid						
Yes	210	34.98	3,390	32.84	3,590	32.97
No	280	65.02	5,080	67.16	5,360	67.03
Receipt of Pell Grant						
Yes	120	18.87	1,910	17.06	2,030	17.17
No	370	81.13	6,560	82.94	6,930	82.83
Receipt of Stafford Loan						
Yes	160	27.91	2,590	25.13	2,750	25.30
No	330	72.09	5,870	74.87	6,200	74.70
Receipt of state aid						
Yes	90	12.77	1,320	12.61	1,400	12.62
No	400	87.23	7,150	87.39	7,550	87.38
Receipt of institution aid						
Yes	120	16.94	2,230	22.50*	2,350	22.16
No	370	83.06	6,230	77.50*	6,600	77.84

See notes at end of table.

Table 33. Comparison of B&B:93/03 converted refusals and other respondents—Continued

Demographic characteristics	Converted refusals		Other respondents		Total respondents	
	Sample size	Percent estimate[1]	Sample size	Percent estimate[1]	Sample size	Percent estimate[1]
Receipt of any aid						
Yes	270	48.03	4,640	49.01	4,910	48.95
No	220	51.97	3,830	50.99	4,050	51.05
Prior respondent						
Respondent to either B&B:93/94 or B&B:93/97	470	89.48	8,390	96.75*	8,860	96.30
Nonrespondent to both surveys	20	10.52	90	3.25*	110	3.70
Parents' income (for dependent students)						
Less than $10,000	10	0.84	180	1.98*	190	1.91
10,000–19,999	20	2.76	320	3.11	330	3.09
20,000–29,999	30	5.35	480	5.10	510	5.11
30,000–39,999	40	6.03	530	6.00	560	6.00
40,000–49,999	40	9.70	630	7.22	670	7.38
50,000–59,999	50	12.40	870	11.40	920	11.46
60,000–69,999	30	7.04	650	8.23	680	8.16
70,000–79,999	20	4.84	400	4.80	410	4.80
80,000–99,999	20	2.89	460	5.62*	480	5.45
100,000 or more	30	6.49	620	7.58	650	7.51
Student's income (for independent students)						
Less than $5,000	40	7.30	650	6.85	680	6.88
5,000–9,999	40	7.05	650	6.75	690	6.77
10,000–19,999	30	4.92	690	8.12*	720	7.92
20,000–29,999	30	6.84	410	5.26	440	5.36
30,000–49,999	40	11.80	530	7.59	570	7.85
50,000 or more	20	3.76	280	4.40	300	4.36
Telephone numbers available						
0 or 1	20	5.29	380	4.77	400	4.80
2	70	14.92	1,160	14.39	1,230	14.42
3	130	23.92	1,980	23.33	2,110	23.37
4	100	20.52	2,000	23.13	2,110	22.97
5	90	21.92	1,880	21.56	1,980	21.58
6	40	7.88	760	8.83	800	8.77
7 or more	30	5.57	320	3.99	360	4.09
Number times answering machine encountered						
0	#	12.39	2,830	33.55*	2,880	32.25
1	20	5.25	840	9.51*	860	9.24
More than 1	420	82.37	4,810	56.94*	5,230	58.51
In field cluster						
Yes	220	46.61	3,790	46.75	4,000	46.74
No	270	53.39	4,700	53.25	4,970	53.26

#Rounds to zero.
*Difference between respondents and nonrespondents is significant at the $p < 0.05/(c-1)$ level, where c is the number of categories within the primary variable.
[1]The weight used is BB03_W1U (unrounded initial weight for B&B:03).
[2]Due to small cell sizes, private, not-for-profit and private, for-profit institutions were combined.
NOTE: Characteristics are from base-year data (NPSAS:93) with the exceptions of citizenship (B&B:93/97), phone numbers available, times answering machine encountered, and whether in field cluster (B&B:93/03). Missing values are excluded from the table. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

Table 34 compares the demographic distributions of those who responded early (July 31, 2003, or earlier) with those who responded later (August 1, 2003, or later). This analysis was performed because it is likely that those who responded later in the interview period may be similar to those who did not respond. This analysis shows that the distributions of early and late responders are similar for many of the demographic characteristics. However, the distributions of race/ethnicity, gender, citizenship status, student stratum, whether the student applied for aid in the base year, receipt of any aid in the base year, whether the student responded to prior rounds of data collection, and locator information (number of telephone numbers available, number of times an answering machine was encountered, and whether the student lived in a field cluster) are significantly different for early versus late respondents.

The nonresponse bias was estimated for variables known for both respondents and nonrespondents. The bias in an estimated mean based on respondents, $\overline{y}_R$, will be estimated as the difference between this mean and the target parameter, B, being estimated (i.e., the mean that would be estimated if a complete census of the target population were conducted). This bias can be expressed as follows:

$$B(\overline{y}_R) = \overline{y}_r - \pi .$$

The estimated mean based on nonrespondents, $\overline{y}_{NR}$, can be computed using data for the particular variable for which the data for most of the nonrespondents were available. π can be estimated as follows:

$$\hat{\pi} = (1-\eta)\overline{y}_R + \eta\overline{y}_{NR}$$

where η is the weighted unit nonresponse rate. Therefore, the bias can be estimated as follows:

$$\hat{B}(\overline{y}_R) = \overline{y}_R - \hat{\pi}$$

or equivalently

$$\hat{B}(\overline{y}_R) = \eta(\overline{y}_R - \overline{y}_{NR}) .$$

This formula shows that the estimate of the nonresponse bias is the difference between the mean for respondents and nonrespondents multiplied by the weighted nonresponse rate.

The variance of the bias is then computed as follows:

$$\text{var}(\hat{B}) = \eta^2 \text{ var}(\overline{y}_R - \overline{y}_{NR})$$

where $\overline{y}_R$ and $\overline{y}_{NR}$ are the estimates using the original weights and $\text{var}(\overline{y}_R - \overline{y}_{NR})$ was estimated using Taylor series linearization (taking into account the covariance between $\overline{y}_R$ and $\overline{y}_{NR}$). A *t* test was used to determine which variables have significant nonresponse bias at the 5 percent level.

Table 34. Comparison of B&B:93/03 late respondents and early respondents

Demographic characteristics	Late respondents		Early respondents		Total respondents	
	Sample size	Percent estimate[1]	Sample size	Percent estimate[1]	Sample size	Percent estimate[1]
Age						
21 or younger	450	26.03	1,980	25.67	2,430	25.74
22	490	29.47	1,960	26.68	2,450	27.26
23	250	14.80	970	13.42	1,230	13.70
24 to 27	240	13.10	1,050	14.53	1,290	14.24
28 or older	260	16.60	1,300	19.70	1,560	19.06
Race/ethnicity						
White, non-Hispanic	1,350	80.54	6,310	86.65*	7,660	85.39
Black or African American, non-Hispanic	120	6.58	380	4.97	500	5.30
Hispanic	120	7.48	310	4.41*	440	5.04
Asian/Native Hawaiian/Other Pacific Islander	80	4.75	220	3.41	300	3.68
American Indian/Alaska Native	10	0.66	40	0.56	60	0.58
Gender						
Male	770	48.41	3,050	43.35*	3,820	44.39
Female	920	51.59	4,230	56.65*	5,150	55.61
U.S. citizenship						
Yes	1,660	96.71	7,210	98.67*	8,870	98.27
No	30	3.29	70	1.33*	100	1.73
Attendance status						
Full time/full year: 1 institution	810	47.76	3,330	45.06	4,140	45.62
Full time/full year: more than 1 institution	40	0.84	160	0.91	190	0.90
Full-time/part year	280	18.01	1,280	17.21	1,560	17.38
Part time/full year: 1 institution	240	15.59	1,100	16.91	1,350	16.64
Part time/full year: more than 1 institution	20	0.72	110	0.90	130	0.86
Part time/part year	280	17.09	1,270	19.00	1,550	18.61
Institution control						
Public	1,090	66.66	4,760	67.90	5,850	67.64
Private, not-for-profit	590	31.11	2,470	31.12	3,050	31.12
Private, for-profit	20	2.23	50	0.98	70	1.23
Institution region						
New England	120	6.43	530	7.17	650	7.02
Mid East	320	17.95	1,150	15.23	1,470	15.79
Great Lakes	270	18.89	1,200	17.39	1,470	17.70
Plains	110	7.06	680	10.01*	790	9.40
Southeast	410	22.47	1,810	23.72	2,210	23.46
Southwest	190	11.91	810	11.51	1,000	11.59
Rocky Mountains	50	2.27	260	3.39	310	3.16
Far West	200	11.80	780	11.15	990	11.28
Outlying areas	30	1.22	40	0.43	70	0.59
Type of institution/enrollment category						
Public						
Fewer than 2,500	10	0.70	120	0.92	140	0.87
2,500–4,999	80	3.27	340	3.72	410	3.63

See notes at end of table.

Table 34. Comparison of B&B:93/03 late respondents and early respondents—Continued

Demographic characteristics	Late respondents		Early respondents		Total respondents	
	Sample size	Percent estimate[1]	Sample size	Percent estimate[1]	Sample size	Percent estimate[1]
Type of institution/enrollment category—Continued						
Public—Continued						
5,000–9,999	170	9.86	730	9.73	900	9.76
10,000–19,999	260	16.96	1,120	17.22	1,380	17.17
20,000 or more	570	35.87	2,440	36.39	3,010	36.28
Private, not-for-profit						
Fewer than 1,000	40	1.60	180	1.75	220	1.72
1,000–2,499	170	8.86	790	8.76	970	8.78
2,500–4,999	150	7.47	560	6.97	710	7.07
5,000–9,999	120	7.13	460	6.66	580	6.76
10,000 or more	110	6.05	470	6.90	580	6.72
Private, for-profit						
Fewer than 1,000	10	1.71	20	0.51	30	0.76
1,000 or more	10	0.52	30	0.47	30	0.48
B&B institution stratum[2]						
Public 4-year first-professional high education	50	3.24	310	4.52	360	4.26
Public 4-year first-professional low education	420	25.74	1,840	26.03	2,250	25.97
Private 4-year first-professional high education	130	8.36	540	8.35	660	8.35
Private 4-year first-professional low education	40	1.92	210	2.43	250	2.32
Public 4-year doctor's high education	70	5.25	280	4.58	350	4.72
Public 4-year doctor's low education	140	9.02	620	9.15	760	9.12
Private 4-year doctor's high education	30	1.37	90	1.06	110	1.13
Private 4-year doctor's low education	30	1.18	100	1.07	130	1.09
Public 4-year master's high education	60	3.59	240	3.49	300	3.51
Public 4-year master's low education	320	18.38	1,280	17.61	1,600	17.77
Private 4-year master's high education	20	0.64	120	0.67	140	0.67
Private 4-year master's low education	200	10.34	820	10.68	1,020	10.61
Public 4-year bachelor's high education	20	0.43	90	0.55	110	0.53
Public 4-year bachelor's low education	20	1.01	100	1.96	120	1.77
Private 4-year bachelor's high education	10	0.44	90	0.89	100	0.80
Private 4-year bachelor's low education	140	9.09	560	6.94	700	7.38
B&B student stratum						
Other students: Combined cell	110	19.52	370	16.64	480	17.23
Business majors	160	17.84	570	16.03	720	16.40
Nonbusiness majors	1,430	62.64	6,340	67.33*	7,770	66.37
Applied for aid						
Yes	930	49.19	4,160	53.18*	5,080	52.36
No	720	50.81	2,960	46.82*	3,680	47.64
Receipt of federal aid						
Yes	680	32.29	2,910	33.15	3,590	32.97
No	1,010	67.71	4,360	66.85	5,360	67.03
Receipt of Pell Grant						
Yes	400	16.64	1,630	17.3	2,030	17.17
No	1,300	83.36	5,630	82.6	6,930	82.83

See notes at end of table.

Table 34. Comparison of B&B:93/03 late respondents and early respondents—Continued

Demographic characteristics	Late respondents		Early respondents		Total respondents	
	Sample size	Percent estimate[1]	Sample size	Percent estimate[1]	Sample size	Percent estimate[1]
Receipt of Stafford Loan						
Yes	530	24.95	2,230	25.39	2,750	25.30
No	1,160	75.05	5,040	74.61	6,200	74.70
Receipt of state aid						
Yes	270	11.89	1,130	12.81	1,400	12.62
No	1,420	88.11	6,130	87.19	7,550	87.38
Receipt of institution aid						
Yes	430	21.69	1,920	22.28	2,350	22.16
No	1,260	78.31	5,340	77.72	6,600	77.84
Receipt of any aid						
Yes	900	45.40	4,010	49.87*	4,910	48.95
No	790	54.60	3,250	50.13*	4,050	51.05
Prior respondent						
Respondent to either B&B:93/94 or B&B:93/97	1,640	92.01	7,220	97.41*	8,860	96.30
Nonrespondent to both surveys	50	7.99	60	2.59*	110	3.70
Parents' income (for dependent students)						
Less than $10,000	50	2.44	140	1.78	190	1.91
10,000–19,999	60	2.74	270	3.18	330	3.09
20,000–29,999	110	6.00	400	4.88	510	5.11
30,000–39,999	100	5.85	460	6.04	560	6.00
40,000–49,999	130	7.55	540	7.33	670	7.38
50,000–59,999	200	13.85	720	10.84	920	11.46
60,000–69,999	110	7.45	560	8.34	680	8.16
70,000–79,999	90	6.24	330	4.43	410	4.80
80,000–99,999	80	5.26	400	5.50	480	5.45
100,000 or more	130	7.77	520	7.45	650	7.51
Students' income (for independent students)						
Less than $5,000	130	5.85	560	7.14	680	6.88
5,000–9,999	130	6.84	560	6.75	690	6.77
10,000–19,999	120	6.47	600	8.30	720	7.92
20,000–29,999	80	4.98	370	5.45	440	5.36
30,000–49,999	100	7.35	470	7.98	570	7.85
50,000 or more	40	3.37	250	4.62	300	4.36
Telephone numbers available						
0 or 1	70	5.06	330	4.73	400	4.80
2	200	13.50	1,030	14.66	1,230	14.42
3	340	19.94	1,760	24.26*	2,110	23.37
4	380	21.59	1,720	23.33	2,110	22.97
5	330	19.69	1,640	22.08	1,980	21.58
6	220	12.05	580	7.92*	800	8.77
7 or more	140	8.19	210	3.02*	360	4.09

See notes at end of table.

Table 34. Comparison of B&B:93/03 late respondents and early respondents—Continued

Demographic characteristics	Late respondents		Early respondents		Total respondents	
	Sample size	Percent estimate[1]	Sample size	Percent estimate[1]	Sample size	Percent estimate[1]
Number times answering machine encountered						
0	140	8.65	2,740	38.36*	2,880	32.25
1	60	3.57	800	10.71*	860	9.24
More than 1	1,490	87.78	3,740	50.93*	5,230	58.51
In field cluster						
Yes	980	60.66	3,030	43.13*	4,000	46.74
No	710	39.34	4,250	56.87*	4,970	53.26

* Difference between respondents and nonrespondents is significant at the $p < 0.05/(c-1)$ level, where c is the number of categories within the primary variable.
[1] The weight used is BB03_W1U (unrounded initial weight for B&B:93/03).
[2]Due to small cell sizes, private, not-for-profit and private, for-profit institutions were combined.
NOTE: Characteristics are from base-year data (NPSAS:93) with the exceptions of citizenship (B&B:93/97), phone numbers available, times answering machine encountered, and whether in field cluster (B&B:93/03). Missing values are excluded from the table. Detail may not sum to totals because of rounding. Late respondents are those who respondented after August 1, 2003.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

The first set of columns in table 35 shows the estimated bias before weighting for characteristics available for most responding and nonresponding students. The bias of several variables is statistically significant, such as race/ethnicity, citizenship, control and region of the base-year institution, whether the student applied for aid in the base year, whether aid was received, and whether the student responded to one of the prior rounds of the survey. However, the bias is small for some of these variables.

The second set of columns in table 36 shows the estimated bias after the nonresponse weight adjustments for variables available for most responding and nonresponding students. Some variables have zero bias after the nonresponse adjustments, and the bias is not significantly different from zero for the remaining variables.

Weight adjustments are typically used to reduce bias due to unit nonresponse, and the results in tables 32–35 show that these adjustments are definitely important for reducing the potential for nonresponse bias due to the differences between respondents and nonrespondents. The nonresponse models incorporated the survey stratification variables, variables identified during the CHAID analysis, and other variables that were thought to be predictive of nonresponse (i.e., the variables in the rows of tables 32–35). The three steps of nonresponse adjustment:

- inability to locate the student;
- refusal to be interviewed; and
- other noninterview

were used to adjust for the potential bias resulting from the three different types of nonresponse. Section 6.1 of this chapter gives details of the weighting.

Table 35. Nonresponse bias before and after weight adjustments for selected variables: 2003

Demographic characteristics	Unweighted respondents	Unweighted non-respondents	Before nonresponse adjustment: Respondent percentage, original weights	Before nonresponse adjustment: Non-respondent percentage, original weights	Before nonresponse adjustment: Estimated bias	Before nonresponse adjustment: Percent relative bias	After nonresponse weight adjustment: Percentage, final adjusted weights	After nonresponse weight adjustment: Estimated bias	After nonresponse weight adjustment: Percent relative bias
Age									
21 or younger	2,430	320	25.74	22.19	0.59	2.34	25.15	-0.01	-0.03
22	2,450	360	27.26	26.08	0.20	0.72	27.09	0.03	0.10
23	1,230	220	13.70	13.84	-0.02	-0.17	13.71	-0.01	-0.09
24 to 27	1,290	220	14.24	16.62	-0.39	-2.69	14.62	-0.01	-0.04
28 or older	1,560	310	19.06	21.28	-0.37	-1.88	19.43	#	-0.02
Race/Ethnicity									
White, non-Hispanic	7,660	1,100	85.39	75.19	1.68	2.01*	83.73	0.02	0.03
Black or African American, non-Hispanic	500	120	5.30	9.74	-0.73	-12.15*	6.03	#	-0.08
Hispanic	440	100	5.04	5.89	-0.14	-2.69	5.16	-0.02	-0.36
Asian/Native Hawaiian/Other Pacific Islander	300	100	3.68	8.70	-0.83	-18.34*	4.51	#	0.04
American Indian/Alaska Native	60	10	0.58	0.47	0.02	3.21	0.56	#	-0.06
Gender									
Male	3,820	670	44.39	48.21	-0.63	-1.40	45.01	-0.01	-0.03
Female	5,150	760	55.61	51.79	0.63	1.15	54.99	0.01	0.02
U.S. citizenship									
Yes	8,870	1,370	98.27	93.09	0.86	0.88*	97.42	#	#
No	100	50	1.73	6.91	-0.86	-33.10*	2.58	#	-0.01
Attendance status									
Full time/full year: 1 institution	4,140	620	45.62	41.05	0.75	1.68	44.89	0.03	0.06
Full time/full year: more than 1 institution	190	30	0.90	1.37	-0.08	-7.99	0.97	#	-0.13
Full time/part year	1,560	270	17.38	19.47	-0.35	-1.95	17.66	-0.07	-0.37
Part time/full year: 1 institution	1,350	220	16.64	17.02	-0.06	-0.38	16.72	0.02	0.10
Part time/full year: more than 1 institution	130	20	0.86	0.85	#	0.33	0.86	#	-0.02
Part time/part year	1,550	250	18.61	20.25	-0.27	-1.43	18.90	0.02	0.12
Institution control									
Public	5,850	910	67.64	62.33	0.88	1.32*	66.76	#	#
Private, not-for-profit	3,050	500	31.12	34.43	-0.55	-1.73	31.67	#	#
Private, for-profit	70	20	1.23	3.25	-0.33	-21.26	1.57	#	#
Institution region									
New England	650	100	7.02	7.12	-0.02	-0.23	7.04	#	0.06
Mid East	1,470	280	15.79	21.72	-0.98	-5.86*	16.77	#	#
Great Lakes	1,470	230	17.70	15.97	0.29	1.65	17.41	-0.01	-0.04
Plains	790	80	9.40	6.61	0.46	5.16	8.94	0.01	0.07

See notes at end of table.

Table 35. Nonresponse bias before and after weight adjustments for selected variables: 2003—Continued

Demographic characteristics	Unweighted respondents	Unweighted non-respondents	Before nonresponse adjustment: Respondent percentage, original weights	Before nonresponse adjustment: Nonrespondent percentage, original weights	Before nonresponse adjustment: Estimated bias	Before nonresponse adjustment: Percent relative bias	After nonresponse weight adjustment: Percentage, final adjusted weights	After nonresponse weight adjustment: Estimated bias	After nonresponse weight adjustment: Percent relative bias
Institution region—Continued									
Southeast	2,210	380	23.46	23.36	0.02	0.07	23.44	#	-0.01
Southwest	1,000	150	11.59	9.02	0.43	3.81	11.16	#	-0.03
Rocky Mountains	310	40	3.16	3.19	-0.01	-0.16	3.16	#	-0.10
Far West	990	160	11.28	12.47	-0.20	-1.72	11.49	0.01	0.07
Outlying areas	70	20	0.59	0.54	0.01	1.58	0.58	#	-0.47
Type of institution/enrollment category									
Public									
Fewer than 2,500	140	30	0.87	1.26	-0.06	-6.84	0.93	#	-0.02
2,500–4,999	410	70	3.63	4.10	-0.08	-2.13	3.70	#	-0.04
5,000–9,999	900	150	9.76	9.48	0.05	0.47	9.71	#	0.01
10,000–19,999	1,380	210	17.17	16.26	0.15	0.89	17.01	#	-0.02
20,000 or more	3,010	450	36.28	31.29	0.83	2.34	35.45	#	-0.01
Private, not-for-profit									
Fewer than 1,000	220	30	1.72	2.31	-0.10	-5.38	1.82	#	-0.02
1,000–2,499	970	150	8.78	7.90	0.15	1.68	8.63	#	-0.05
2,500–4,999	710	100	7.07	6.69	0.06	0.91	7.01	#	#
5,000–9,999	580	100	6.76	7.50	-0.12	-1.78	6.89	0.01	0.08
10,000 or more	580	120	6.72	9.96	-0.54	-7.38	7.27	0.01	0.08
Private, for-profit									
Fewer than 1,000	30	10	0.76	2.19	-0.24	-23.74	1.00	#	-0.02
1,000 or more	30	10	0.48	1.07	-0.10	-16.97	0.58	#	-0.02
B&B institution stratum[1]									
Public 4-year first-professional high education	360	50	4.26	3.17	0.18	4.42	4.08	#	#
Public 4-year first-professional low education	2,250	320	25.97	21.51	0.74	2.93	25.23	#	#
Private 4-year first-professional high education	660	140	8.35	11.75	-0.56	-6.32	8.91	#	#
Private 4-year first-professional low education	250	40	2.32	3.25	-0.15	-6.18	2.48	#	#
Public 4-year doctor's high education	350	50	4.72	3.90	0.14	2.95	4.58	#	#
Public 4-year doctor's low education	760	120	9.12	7.62	0.25	2.80	8.87	#	#
Private 4-year doctor's high education	110	30	1.13	2.15	-0.17	-13.12	1.30	#	#
Private 4-year doctor's low education	130	20	1.09	0.74	0.06	5.58	1.03	#	#
Public 4-year master's high education	300	50	3.51	3.45	0.01	0.30	3.50	#	#
Public 4-year master's low education	1,600	280	17.77	20.34	-0.43	-2.34	18.19	#	#
Private 4-year master's high education	140	20	0.67	0.47	0.03	5.09	0.64	#	#
Private 4-year master's low education	1,020	160	10.61	10.25	0.06	0.56	10.55	#	#
Public 4-year bachelor's high education	110	10	0.53	0.29	0.04	8.06	0.49	#	#
Public 4-year bachelor's low education	120	20	1.77	2.05	-0.05	-2.62	1.81	#	#
Private 4-year bachelor's high education	100	10	0.80	0.65	0.03	3.32	0.78	#	#
Private 4-year bachelor's low education	700	110	7.38	8.40	-0.17	-2.24	7.55	#	#

See notes at end of table.

Table 35. Nonresponse bias before and after weight adjustments for selected variables: 2003—Continued

Demographic characteristics	Unweighted respondents	Unweighted non-respondents	Before nonresponse adjustment: Respondent percentage, original weights	Before nonresponse adjustment: Non-respondent percentage, original weights	Before nonresponse adjustment: Estimated bias	Before nonresponse adjustment: Percent relative bias	After nonresponse weight adjustment: Percentage, final adjusted weights	After nonresponse weight adjustment: Estimated bias	After nonresponse weight adjustment: Percent relative bias
B&B student stratum									
Other students: Combined cell	480	90	17.23	18.94	-0.28	-1.61	17.51	#	#
Business majors	720	140	16.40	16.18	0.04	0.22	16.37	#	#
Non business majors	7,770	1,210	66.37	64.88	0.25	0.37	66.12	#	#
Applied for aid									
Yes	5,080	770	52.36	47.71	0.76	1.47*	51.60	-0.01	-0.02
No	3,680	610	47.64	52.29	-0.76	-1.57*	48.40	0.01	0.02
Receipt of federal aid									
Yes	3,590	570	32.97	32.59	0.06	0.19	32.90	-0.01	-0.02
No	5,360	860	67.03	67.41	-0.06	-0.10	67.10	0.01	0.01
Receipt of Pell Grant									
Yes	2,030	340	17.17	18.76	-0.26	-1.50	17.42	-0.01	-0.06
No	6,930	1,080	82.83	81.24	0.26	0.32	82.58	0.01	0.01
Receipt of Stafford Loan									
Yes	2,750	430	25.30	24.64	0.11	0.43	25.19	#	-0.01
No	6,200	1,000	74.70	75.36	-0.11	-0.15	74.81	#	#
Receipt of state aid									
Yes	1,400	210	12.62	10.75	0.31	2.52	12.30	-0.01	-0.05
No	7,550	1,220	87.38	89.25	-0.31	-0.35	87.70	0.01	0.01
Receipt of institution aid									
Yes	2,350	300	22.16	17.80	0.72	3.36*	21.63	0.19	0.88
No	6,600	1,120	77.84	82.20	-0.72	-0.92*	78.37	-0.19	-0.24
Receipt of any aid									
Yes	4,910	740	48.95	43.97	0.83	1.72*	48.12	#	-0.01
No	4,050	690	51.05	56.03	-0.83	-1.59*	51.88	#	#
Prior respondent									
Respondent to either B&B:93/94 or B&B:93/97	8,860	1,320	96.30	84.37	1.98	2.10*	94.32	#	#
Nonrespondent to both surveys	110	110	3.70	15.63	-1.98	-34.85*	5.68	#	#
Parents' income (for dependent students)									
Less than $10,000	190	40	1.91	1.89	#	0.24	1.90	-0.01	-0.28
10,000–19,999	330	50	3.09	2.77	0.05	1.69	3.01	-0.03	-0.83
20,000–29,999	510	80	5.11	4.61	0.08	1.65	5.04	0.01	0.26

See notes at end of table.

Table 35. Nonresponse bias before and after weight adjustments for selected variables: 2003—Continued

Demographic characteristics	Unweighted respondents	Unweighted non-respondents	Before nonresponse adjustment: Respondent percentage, original weights	Before nonresponse adjustment: Non-respondent percentage, original weights	Before nonresponse adjustment: Estimated bias	Before nonresponse adjustment: Percent relative bias	After nonresponse weight adjustment: Percentage, final adjusted weights	After nonresponse weight adjustment: Estimated bias	After nonresponse weight adjustment: Percent relative bias
Parents' income (for dependent students) —Continued									
30,000–39,999	560	80	6.00	5.75	0.04	0.68	5.97	0.01	0.20
40,000–49,999	670	80	7.38	6.65	0.12	1.64	7.25	#	-0.03
50,000–59,999	920	170	11.46	12.30	-0.14	-1.19	11.64	0.04	0.38
60,000–69,999	680	100	8.16	7.00	0.19	2.39	7.99	0.02	0.26
70,000–79,999	410	60	4.80	4.50	0.05	1.05	4.81	0.06	1.33
80,000–99,999	480	60	5.45	4.76	0.11	2.14	5.33	#	-0.06
100,000 or more	650	80	7.51	6.40	0.18	2.49	7.34	0.01	0.08
Students' income (for independent students)									
Less than $5,000	680	130	6.88	9.00	-0.35	-4.82	7.12	-0.10	-1.45
5,000–9,999	690	100	6.77	6.61	0.03	0.38	6.75	#	0.07
10,000–19,999	720	150	7.92	10.72	-0.46	-5.48	8.37	-0.01	-0.13
20,000–29,999	440	80	5.36	5.59	-0.04	-0.70	5.39	-0.01	-0.15
30,000–49,999	570	90	7.85	7.16	0.11	1.46	7.74	#	-0.01
50,000 or more	300	50	4.36	4.30	0.01	0.23	4.35	#	-0.07
Telephone numbers available									
0 or 1	400	90	4.80	8.36	-0.59	-10.95*	5.39	#	#
2	1,230	230	14.42	18.58	-0.69	-4.56*	15.11	#	#
3	2,110	320	23.37	22.70	0.11	0.47	23.26	#	#
4	2,110	300	22.97	19.33	0.60	2.69	22.37	#	#
5	1,980	250	21.58	16.24	0.89	4.28*	20.70	#	#
6	800	140	8.77	8.44	0.05	0.62	8.71	#	#
7 or more	360	90	4.09	6.33	-0.37	-8.34	4.46	#	#
Number times answering machine encountered									
0	2,880	220	32.25	16.98	2.53	8.51*	29.72	#	#
1	860	90	9.24	6.14	0.51	5.90*	8.73	#	#
More than 1	5,230	1,110	58.51	76.88	-3.04	-4.94*	61.55	#	#
In field cluster									
Yes	4,000	700	46.74	53.01	-1.04	-2.17*	47.78	#	#
No	4,970	730	53.26	46.99	1.04	1.99*	52.22	#	#

Rounds to zero.
* Difference between respondents and nonrespondents is significant at the $p < 0.05/(c-1)$ level, where c is the number of categories within the primary variable.
Original weight is BB03_W1U (unrounded initial weight for B&B:03). Final adjusted weight is BNBWT3 (rounded cross-sectional analysis weight).
[1]Due to small cell sizes, private, not-for-profit and private, for-profit institutions were combined.
NOTE: Characteristics are from base-year data (NPSAS:93) with the exceptions of citizenship (B&B:93/97), and phone numbers available, times answering machine encountered, and whether in field cluster (B&B:93/03). Missing values are excluded from the table. Detail may not sum to totals due to rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03.)

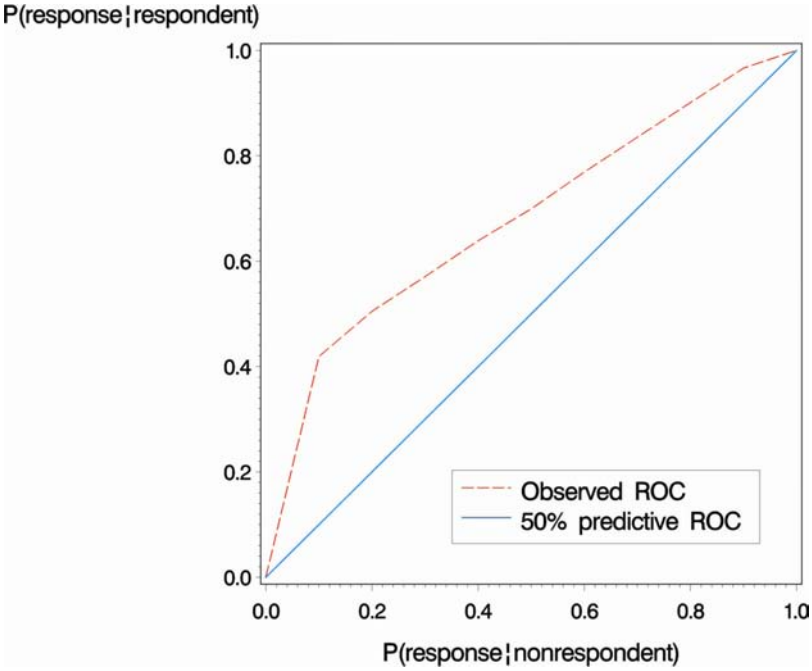
6.3.2.2 ROC curve

As described in section 6.1, three nonresponse adjustment models were used for computing the final cross-sectional analysis weights for B&B:93/03. To assess the overall predictive ability of the combined models, a Receiver Operating Characteristics (ROC) curve was used. A point on an ROC curve is constructed by considering a given predicted probability as a cutoff point for deciding whether a person is a respondent or a nonrespondent. For a given cutoff, a point on the ROC curve is obtained by plotting the proportion of respondents with a predicted probability greater than the cutoff (i.e., true positives) versus the proportion of nonrespondents with a predicted probability greater than the cutoff (i.e., false positives). The points on the ROC curve are then obtained by computing the proportion of true and false positives for the entire range of possible cutoffs.

The area under an ROC curve measures the probability that the fitted model will correctly classify two randomly chosen individuals—one of which is a true respondent and the other a true nonrespondent—where the individual with the higher predicted probability of response is classified as a respondent. An area of 0.5 under an ROC curve indicates that a correct classification is a 50:50 proposition, with the model providing no predictive benefit. An area of 1 indicates that the true respondent always has the higher predicted probability of response and, thus, that the model always classifies the two individuals correctly. The area under the ROC curve in figure 9 is 0.69, so 69 percent of the time the predicted probabilities give the correct classification.

All of the students in the B&B:93/03 sample were used for constructing this ROC curve. The student's predicted probability of response was calculated as the product of the predicted probabilities obtained from the three GEM models described in section 6.1: P(located), P(nonrefusal for located students), and P(response for located students who did not refuse). These probabilities are the inverse of the adjustment factors from the GEM models. Since only located students were included in the nonrefusal model and only nonrefusals were included in the final response model, the predicted probabilities will not be directly available for students who were not located or for students who refused. The mean of the predicted probabilities was used for students who were in the models for the probabilities that were not directly available.

Figure 9. ROC curve for overall response propensity: 2003



NOTE: The area under the curve = 0.69.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

6.4 Response Rates

6.4.1 Overall Response Rates

The overall B&B:93/03 study response rate is an estimate of the proportion of the study population directly represented by the study respondents. Because the B&B:93/03 study includes a subsample of nonrespondents to NPSAS:93, B&B:93/94, and B&B:93/97, the overall study response rate is the product of the NPSAS:93 institution-level response rate times the B&B:93/03 student-level response rate. Therefore, the overall B&B:93/03 study response rates can be estimated directly only for domains defined by institutional characteristics.

Both weighted and unweighted overall study response rates are shown in table 36, along with their institution and student response rate components. The institution-level response rates shown in this table are the percentage of institutions that provided sufficient data to select the NPSAS student-level sample; these rates were obtained from the NPSAS:93 Methodology Report (Loft et al. 1995, table 2.7). Only the weighted response rates can be interpreted as estimates of the proportion of the B&B study population that are directly represented by the study respondents. Table 36 shows that the student response rate is 83 percent and that approximately 74 percent of the B&B study population is represented by the respondents. The rate of population coverage varies by type of institution: the rate is higher for public institutions than for private institutions, and is higher for institutions offering a master's or doctoral degree than for those offering bachelor's or less, or first-professional degree.

Each weighted student response rate was calculated as the weighted number of respondents divided by the weighted number of eligible students. The weight used in these calculations was the B&B:93/94 base weight that has been adjusted for subsampling the B&B:93/97 nonrespondents; this is the weight variable BB03_W1. Each overall study response rate was calculated as the product of the NPSAS:93 institutional response rate times the student response rate.

The overall response rates for B&B:93/03 are presented in table 37 by prior response status. The weighted response rates are highest for those students who were interviewed in B&B:93/97. Among those interviewed in all three of the previous surveys, the weighted response rate was 77 percent. Among those interviewed in only B&B:93/97 and B&B:93/94 or in only B&B:93/97 and NPSAS:93, the weighted response rates were 75 percent and 69 percent. The rates were much lower for those who were not interviewed in B&B:93/97 (53 percent or less).

6.4.2 Bias Due to Item Nonresponse

This section looks at the bias associated with item nonresponse for those questionnaire items with 15 percent or more item nonresponse. Table 38 presents those items that have 15 percent or more item nonresponse and also had at least 50 potential respondents, along with the nonresponse rates. Potential respondents include those who responded to the item, those who should have responded to the item but did not, and those who did not respond to the gate question for an item. The nonresponse rates presented in table 38 include both of the latter two groups. For each of these items, tables in appendix I compare the demographic characteristics of respondents and nonrespondents with respect to

- age;
- race/ethnicity;
- gender;
- control of NPSAS:93 institution; and
- region.

The bias and the statistical significance of the bias were also estimated, using the formulas and methodology described in section 6.3.2. The final cross-sectional weight, BNBWT3, was used for the calculations. While some variables do show statistically significant biases, the actual bias is generally very small.

Table 36. Overall B&B:93/03 study response rates, by institutional level, control, and sector

Type of student	Institutions[1]				Students				Overall response rate[2]	
	Number eligible	Number responding	Response rate		Number eligible	Number responding	Response rate			
			Unweighted	Weighted			Unweighted	Weighted	Unweighted	Weighted
Total	1,240	1,100	88.3	88.2	10,400	8,970	86.3	83.4	76.2	73.6
Institutional level										
Bachelor's	610	520	86.0	86.7	1,180	1,030	87.1	82.2	74.9	71.3
Master's	290	270	95.1	98.1	3,570	3,070	85.9	82.6	81.7	81.0
Doctor's	90	80	93.0	94.6	1,570	1,350	85.8	84.9	79.8	80.3
First-professional	260	220	84.8	74.6	4,080	3,530	86.5	83.9	73.4	62.6
Institutional control										
Public	620	580	92.3	96.3	6,760	5,850	86.5	84.5	79.9	81.4
Private, not-for-profit	440	380	87.2	91.3	3,550	3,050	86.0	82.0	75.0	74.9
Private, for profit	180	140	77.5	80.1	90	70	76.1	65.7	59.0	52.6
Institutional sector										
Public, bachelor's	310	280	91.5	96.9	260	220	87.5	83.1	80.0	80.6
Public, master's	150	140	95.3	95.4	2,240	1,910	85.2	81.8	81.2	78.1
Public, doctor's	60	50	92.7	94.2	1,290	1,110	86.2	85.8	79.9	80.8
Public, first-professional	120	100	90.4	91.7	2,980	2,610	87.6	86.1	79.2	78.9
Private, not-for-profit, bachelor's	130	110	85.6	89.5	880	770	88.0	85.3	75.3	76.3
Private, not-for-profit, master's	130	130	94.7	98.5	1,290	1,120	87.0	84.4	82.4	83.1
Private, not-for-profit, doctor's or first-professional	180	150	82.7	71.5	1,380	1,160	83.7	78.4	69.2	56.1
Private, for profit	180	140	77.4	80.1	90	70	76.1	65.7	58.9	52.6

See table 2.7 in the NPSAS:93 methodology report. (Loft et al. 1995)
[2] Calculated as the product of the institutional response rate times the student response rate.
Note: The weight used is BB03_1. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

Table 37. B&B:93/03 response rates, by prior response status

Prior response status	Number eligible	Unweighted			Weighted[1]			Overall response rate[2]	
		Respondents	Non-respondents	Response rate	Respondents	Non-respondents	Response rate	Unweighted	Weighted
Total students	10,400	8,970	1,430	86.3	980,240	194,640	83.4	76.2	73.6
Interviewed in NPSAS:93, B&B:93/94, and B&B:93/97	9,170	8,090	1,080	88.3	847,030	122,100	87.4	77.9	77.1
Interviewed in NPSAS:93 and B&B:93/94 only	130	70	60	51.6	23,340	22,230	51.2	45.6	45.2
Interviewed in B&B:93/94 and B&B:93/97 only	380	320	60	83.6	30,430	5,410	84.9	73.8	74.9
Interviewed in NPSAS:93 and B&B:93/97 only	460	360	110	77.1	39,150	11,250	77.7	68.1	68.5
Interviewed in NPSAS:93 only	200	110	100	53.2	35,430	27,710	56.1	47.0	49.5
Interviewed in B&B:93/94 only	10	10	#	60.0	2,220	1,460	60.3	53.0	53.2
Interviewed in B&B:93/97 only	40	20	20	55.6	1,830	1,760	50.9	49.1	44.9
Not interviewed in prior surveys	10	#	10	23.1	810	2,720	23.0	20.4	20.3

#Rounds to zero.
[1] The weight used in computing the weighted counts of respondents and nonrespondents is B03_W1. This weight was applied to the 10,400 eligible sample members.
[2] The overall response rate is the product of the institution response rate from table 36 times the student response rates. Only those students who responded to the NPSAS:93 CATI interview are counted as NPSAS:93 respondents.
NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

Table 38. Items with 15 percent or more item nonresponse rate and at least 50 persons in the denominator of the rate

Variable name	Description	Persons eligible to answer the item		Persons responding to gate/decision items	
		Percent item nonresponse	Number of persons in the denominator	Percent item nonresponse	Number of persons in the denominator
B3CTAMT2	Certificate 2: time spent	55.27	190	0.00	150
B3CTCOM2	Certificate 2: total hours to complete	55.55	190	0.62	150
B3CTEXP1	Certificate 1: expect to complete	32.18	600	1.80	420
B3CTTIM2	Certificate 2: unit of time	25.88	190	0.00	150
B3GR2EMY	Graduate: date earned degree 2	28.17	170	0.00	140
B3GRRE1	Graduate 1: when plan to return	19.78	140	1.48	110
B3MSTR2	Master's degree type at grad school 2	19.78	170	1.48	130
B3STLG2	Graduate: still working toward degree 2	31.50	120	0.12	50
B3CUROTH	Current job: hours non-primary job	31.50	930	0.00	830
B3LOOK3	Looking for work while unemployed 3	26.56	380	2.00	310
B3LOOK4	Looking for work while unemployed 4	50.94	260	0.00	190
B3LOOK5	Looking for work while unemployed 5	25.88	210	0.00	140
B3LOOK6	Looking for work while unemployed 6	31.50	180	0.00	120
B3LOOK7	Looking for work while unemployed 7	31.50	170	0.00	100
B3NOACA	Declined offer: low salary/benefits	31.50	100	0.00	30
B3NOACB	Declined offer: accepted other job	31.50	100	0.00	30
B3NOACC	Declined offer: other job more pay	31.50	100	0.00	30
B3NOACD	Declined offer: other job interesting	31.50	100	0.00	30
B3NOACE	Declined offer: too far from home	37.88	100	1.98	30
B3NOACF	Declined offer: dangerous/difficult	20.17	100	1.97	30
B3NOACG	Declined offer: not qualified	15.42	100	7.20	30
B3NOACX	Declined offer: other	55.27	100	0.00	30
B3AIDPAY	Time as aide paid or unpaid	55.55	220	0.62	110
B3AIDREG	Took teacher's aide job to go permanent	32.18	220	1.80	110
B3HLPA1	Help new teachers: student discipline	25.88	410	0.00	340
B3HLPB1	Help new teachers: instructional methods	28.17	410	0.00	340
B3HLPC1	Help new teachers: the curriculum	19.78	410	1.48	340

See notes at end of table.

Table 38. Items with 15 percent or more item nonresponse rate and at least 50 persons in the denominator of the rate—Continued

Variable name	Description	Persons eligible to answer the item		Persons responding to gate/decision items	
		Percent item nonresponse	Number of persons in the denominator	Percent item nonresponse	Number of persons in the denominator
B3HLPD1	Help new teachers: adjust	19.11	410	0.25	340
B3INDUCT	First: formal induction program	19.09	410	0.22	340
B3LFTTCH	Non-teaching position planned	22.85	700	28.32	630
B3OFRCHK	Verification of past teaching	65.28	120	0.00	50
B3PSTTYP	Type job last non-substitute	85.41	130	12.26	30
B3SUBLNG	Long-term substitute	22.75	520	6.72	450
B3SUBREG	Took substitute job to go permanent	22.58	520	6.51	450
B3SUPAA	Support: professional development	47.69	250	5.09	150
B3SUPAB	Support: curricular activities/materials	47.69	250	5.09	150
B3SUPAC	Support: technical assistance	47.69	250	5.09	150
B3SUPAX	Support: other	47.69	250	5.09	150
B3TOINC1	Other income from school district	28.59	580	8.29	480
B3TOOIN1	Non-school income	28.87	580	8.50	480
B3TS2BMY	Teaching beginning date 2	26.24	810	23.99	780
B3TS2EMY	Teaching ending date 2	24.87	810	22.58	780
B3TS2TBL	Table containing school 2 (B3TSIZ) details	23.54	810	11.94	730
B3TSALR1	Base annual teaching salary	28.41	580	8.07	480
B3UNF1A	Unprepared: art/drama/music	52.53	220	2.46	120
B3UNF1B	Unprepared: business	52.53	220	2.46	120
B3UNF1C	Unprepared: economics/political systems	52.53	220	2.46	120
B3UNF1D	Unprepared: elementary/early childhood	52.53	220	2.46	120
B3UNF1E	Unprepared: language arts	52.53	220	2.46	120
B3UNF1F	Unprepared: ESL/bilingual	52.53	220	2.46	120
B3UNF1G	Unprepared: foreign languages	52.53	220	2.46	120
B3UNF1H	Unprepared: health/physical education	52.53	220	2.46	120
B3UNF1I	Unprepared: mathematics	52.53	220	2.46	120
B3UNF1J	Unprepared: science	52.53	220	2.46	120
B3UNF1K	Unprepared: secondary education	52.53	220	2.46	120

See notes at end of table.

Table 38. Items with 15 percent or more item nonresponse rate and at least 50 persons in the denominator of the rate—Continued

Variable name	Description	Persons eligible to answer the item		Persons responding to gate/decision items	
		Percent item nonresponse	Number of persons in the denominator	Percent item nonresponse	Number of persons in the denominator
B3UNF1L	Unprepared: special education	52.53	220	2.46	120
B3UNF1M	Unprepared: social studies/history	52.53	220	2.46	120
B3UNF1N	Unprepared: vocational/occupational	52.53	220	2.46	120
B3UNF1O	Unprepared: social sciences	52.53	220	2.46	120
B3UNF1X	Unprepared: other	52.53	220	2.46	120
B3FGVTYP	Type of loan forgiveness program	96.39	1,430	2.27	60
B3FRGVLN	Participate in loan forgiveness program	39.14	3,870	2.34	2,510
B3RPYBY	Year began repaying loans	22.27	2,280	1.98	1,890
B3RPYOY	Year repaid loans	85.64	1,640	9.65	260
B3RPYST	Currently repaying any education loans	17.39	2,740	0.35	2,360
B3RPYTYP	Type of repayment plan	21.88	2,280	1.49	1,890
B3SREPOY	Year that spouse repaid loans	36.54	1,510	20.23	1,210
B3SRPOY	Year that spouse began repaying loans	40.22	1,210	3.49	790
B3SRPST	Spouse currently repaying student loans	33.60	1,380	0.25	960
B3SRPTP	Spouse's type of repayment plan	39.17	1,210	1.79	790
B3CONDIS	Consider to have a disability	32.18	390	1.80	280
B3DSOTA	Impairment: hearing	54.00	390	33.38	280
B3DSOTB	Impairment: visual	54.00	390	33.38	280
B3DSOTC	Impairment: speech	54.00	390	33.38	280
B3DSOTD	Impairment: mobility	54.00	390	33.38	280
B3DSOTE	Impairment: learning disability	54.00	390	33.38	280
B3DSOTF	Impairment: mental	54.00	390	33.38	280
B3DSOTG	Impairment: other	54.00	390	33.38	280

NOTE: The percentages are weighted using the weight variable BNBWT3.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

6.5 Imputation Methodology for the Baccalaureate and Beyond: 2003 Study

Consistent with the statistical standards adopted by the National Center for Education Statistics (NCES) in September 2002, key B&B:93/03 variables to be used in cross-sectional estimates were imputed. The variables identified for imputation were those which appear in the rows and columns of tables presented in the two B&B:93/03 descriptive reports (Bradburn, Nevill, and Forrest Cataldi forthcoming; Henke and Bugarin forthcoming). The imputations were performed in three steps. In the first step, the interview variables were imputed using the procedures described in the next section. Then, using the interview variables, including the newly imputed variable values, the set of derived variables was constructed. In the final step, the derived variables were imputed again, using the procedures described below. Table 39 lists the 26 categorical CATI variables that were imputed. Only one continuous variable was imputed. Income from work in 2002 (B3INC02) had a weighted mean of $50,846 (n = 8,540) prior to imputation and a weighted mean of $50,961 (n = 8,810) after imputation. Similarly, table 40 lists the 16 categorical derived variables that were imputed and table 41 lists the four continuous derived variables that were imputed.

Sequential hot deck imputation, a common procedure for managing item nonresponse, uses respondent data as donors to provide surrogate values for records with missing data. In sequential hot deck imputation, imputation classes are defined, generally consisting of a cross-classification of covariates, and then missing values are replaced sequentially from a single pass through the data within the imputation classes. A related procedure, weighted sequential hot deck imputation, takes into account the unequal probabilities of selection into the original sample by using the sampling weights to specify the expected number of times a particular respondent's answer will be used to replace a missing item. The expected selection frequencies are specified such that, over repeated applications of the algorithm, the expected value of the weighted distribution of the imputed values will equal in expectation, within imputation class, the weighted distribution of the reported answers.

Weighted sequential hot deck imputation was selected for B&B:93/03 in part because it has the advantage of controlling the number of times a respondent record can be used for imputation and gives each respondent record the chance to be selected for use as a hot deck donor. To implement the procedure, imputation classes and sorting variables relevant to each item being imputed were defined. If more than one sorting variable was used, a serpentine sort was performed in which the direction of the sort (ascending or descending) changed each time the value of the previous sorting variable changed. The serpentine sort minimized the change in student characteristics every time one of the sorting variables changed its value.

Imputation classes for the B&B:93/03 interview variables, and some of the derived variables, were developed using a CHAID analysis where only respondent data were modeled. The CHAID segmentation process first divided the data into groups based on categories of the most significant predictor of the item being imputed, and then split each of the groups into smaller subgroups based on the other predictor variables. The CHAID process also merged categories for variables found not to be significantly different. This splitting and merging process continued until no additional statistically significant predictors were found. Imputation classes for B&B:93/03 were then defined from the final CHAID segments.

Table 39. Before and after imputation distributions of categorical variables in the B&B:93/03 interview

Variable	Before imputation		After imputation	
	Sample size	Weighted percent	Sample size	Weighted percent
Marital status (B3MAR)	8,860	100.00	8,970	100.00
Single, never married	1,700	19.41	1,720	19.37
Married	6,110	68.23	6,180	68.23
Cohabiting/living with a partner	400	4.50	410	4.46
Separated	110	1.49	110	1.47
Divorced	510	5.93	520	6.03
Widowed	40	0.45	40	0.44
Community service or volunteer past year (B3COMSRV)	8,880	100.00	8,970	100.00
No, community service in past year	4,620	53.62	4,680	53.85
Yes, community service in past year	4,260	46.38	4,290	46.15
Volunteer: education-related (B3VLTPA)	8,880	100.00	8,970	100.00
No, education-related community service	2,810	31.35	2,830	31.23
Yes, education-related community service	1,450	14.98	1,470	14.92
{Skipped}	4,620	53.67	4,680	53.85
Volunteer: other work with kids (B3VLTPB)	8,880	100.00	8,970	100.00
No, community service work with kids	2,990	32.60	3,010	32.50
Yes, community service work with kids	1,270	13.73	1,280	13.65
{Skipped}	4,620	53.67	4,680	53.85
Volunteer: fundraising (B3VLTPC)	8,880	100.00	8,970	100.00
No, community service fundraising	2,680	28.43	2,700	28.31
Yes, community service fundraising	1,580	17.90	1,600	17.84
{Skipped}	4,620	53.67	4,680	53.85
Volunteer: help for homeless/community (B3VLTPD)	8,880	100.00	8,970	100.00
No, community service help for homeless	3,130	34.10	3,160	33.92
Yes, community service help for homeless	1,130	12.23	1,140	12.23
{Skipped}	4,620	53.67	4,680	53.85
Volunteer: service to the church (B3VLTPE)	8,880	100.00	8,970	100.00
No, community service help to church	2,340	26.21	2,360	26.17
Yes, community service help to church	1,920	20.12	1,940	19.98
{Skipped}	4,620	53.67	4,680	53.85
Volunteer: frequency (B3VLFRQ)	8,870	100.00	8,970	100.00
Daily	80	1.00	80	1.00
Weekly	1,180	12.42	1,190	12.36
Monthly	1,270	13.26	1,280	13.21
Less often	1,420	15.90	1,430	15.87
One time only	310	3.76	310	3.71
{Skipped}	4,620	53.66	4,680	53.85

See notes at end of table.

Table 39. Before and after imputation distributions of categorical variables in the B&B:93/03 interview—Continued

Variable	Before imputation		After imputation	
	Sample size	Weighted percent	Sample size	Weighted percent
Registered to vote (B3VTREG)	8,820	100.00	8,970	100.00
Not registered to vote	610	6.95	670	8.17
Registered to vote	8,210	93.05	8,300	91.83
Voted in the November 2002 National Election (B3VTNEL)	8,180	100.00	8,970	100.00
Did not vote in Nov 02 national election	1,540	18.83	2,230	25.55
Voted in Nov 02 national election	6,640	81.17	6,740	74.45
Political activities past 2 years (B3POLIT)	8,870	100.00	8,970	100.00
No, political activities past 2 years	7,450	85.14	7,540	85.11
Yes, political activities past 2 years	1,420	14.86	1,430	14.89
Make a telephone call (B3TELPN)	8,890	100.00	8,970	100.00
Did not make call to express opinion	7,750	88.30	7,820	88.17
Made call to express opinion	1,140	11.70	1,150	11.83
Undergraduate value: particular major(s) chosen (B3UGVLA)	8,970	100.00	8,970	100.00
Not very important, major chosen	3,680	42.01	3,680	42.01
Very important, major chosen	5,290	57.99	5,290	57.99
Not very important, liberal arts course	5,570	64.08	5,570	64.09
Very important, liberal arts course	3,400	35.92	3,400	35.91
Undergraduate value: professional courses taken (B3UGVLC)	8,970	100.00	8,970	100.00
Not very important, professional course	4,590	50.63	4,600	50.63
Very important, professional course	4,370	49.37	4,370	49.37
Undergraduate value: quality of instruction (B3UGVLD)	8,970	100.00	8,970	100.00
Not very important, quality of instruct	3,390	39.42	3,390	39.42
Very important, quality of instruction	5,580	60.58	5,580	60.58
Undergraduate value: internship and other work (B3UGVLE)	8,970	100.00	8,970	100.00
Not very important, internship	5,080	58.49	5,080	58.49
Very important, internship	3,890	41.51	3,890	41.51
Undergraduate value: none of the above (B3UGVLF)	8,970	100.00	8,970	100.00
Not very important, none of above	8,200	90.59	8,210	90.59
Very important, none of above	760	9.41	760	9.41
Undergraduate preparation: work and career (B3UGPRA)	8,970	100.00	8,970	100.00
Not very important, work and career	1,920	21.76	1,920	21.76
Very important, work and career	7,050	78.24	7,050	78.24

See notes at end of table.

Table 39. Before and after imputation distributions of categorical variables in the B&B:93/03 interview—Continued

Variable	Before imputation		After imputation	
	Sample size	Weighted percent	Sample size	Weighted percent
Undergraduate preparation: further education (B3UGPRB)	8,970	100.00	8,970	100.00
Not very important, further education	3,770	44.12	3,770	44.13
Very important, further education	5,200	55.88	5,200	55.87
Undergraduate preparation: financial security (B3UGPRC)	8,970	100.00	8,970	100.00
Not very important, financial security	3,860	42.66	3,860	42.65
Very important, financial security	5,100	57.34	5,110	57.35
Not very important, none of above	8,250	91.90	8,250	91.89
Very important, none of above	720	8.10	720	8.11
Undergraduate education worth cost (B3UGWRA)	8,940	100.00	8,970	100.00
Undergraduate education not worth cost	920	10.13	930	10.15
Undergraduate education worth cost	8,020	89.87	8,040	89.85
Undergraduate education worth time (B3UGWRB)	8,950	100.00	8,970	100.00
Undergraduate education not worth time	610	7.29	620	7.29
Undergraduate education worth time	8,340	92.71	8,350	92.71
Undergraduate education worth effort (B3UGWRC)	8,900	100.00	8,970	100.00
Undergraduate education not worth effort	400	4.55	400	4.55
Undergraduate education worth effort	8,500	95.45	8,570	95.45
Current employment status (B3CUREMP)	8,960	100.00	8,970	100.00
Not currently employed	400	4.61	400	4.60
Currently employed	7,780	87.22	7,790	87.23
No, waiting for work/temp laid off	80	0.97	80	0.97
No, I am a homemaker	650	6.70	650	6.70
No, I am disabled	50	0.49	50	0.49

NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

Table 40. Before and after imputation distributions of derived categorical variables: 2003

Variable	Before imputation		After imputation	
	Sample size	Weighted percent	Sample size	Weighted percent
Labor force in participation 2003 (B3LFP03)	8,900	100.00	8,970	100.00
Full time, one job	6,130	69.52	6,150	69.12
Part time, one job	730	8.33	770	8.78
Multiple jobs	860	9.29	870	9.33
Unemployed	320	3.87	320	3.85
Out of the labor force	860	8.99	860	8.92
Occupational category (collapsed) (B3OCCAT)	8,950	100.00	8,970	100.00
Educators	1,920	18.70	1,920	18.71
Business and management	2,130	28.55	2,130	28.59
Engineering/architecture	440	4.28	440	4.30
Computer science	410	4.65	410	4.63
Medical professionals	990	9.73	990	9.71
Editors/writers/performers	360	3.98	360	3.96
Human/protective service professionals	790	7.67	790	7.67
Research, scientists, technical	500	5.22	500	5.21
Administrative/clerical/legal support	290	3.55	290	3.54
Mechanics, laborers	180	2.11	180	2.10
Service industries	700	8.56	700	8.57
Other, military	110	1.21	110	1.20
Not applicable	160	1.81	160	1.80
Total number of dependent children in 2003 (B3NUMCH)	8,700	100.00	8,970	100.00
No children	4,130	47.87	4,350	49.07
1	1,830	21.17	1,860	20.69
2	1,910	22.14	1,930	21.61
3	610	6.66	620	6.51
4	170	1.63	170	1.59
5	40	0.36	40	0.36
Over 5	20	0.17	20	0.17
Wrote letter or e-mail to public official 2003 (B3WROTE)	8,890	100.00	8,970	100.00
Did not write	5,910	68.41	5,960	68.34
Wrote letter or email	2,990	31.59	3,010	31.66
Highest degree attained (B3HDG03)	8,790	100.00	8,970	100.00
Bachelor’s degree	6,150	72.28	6,290	73.09
Postbaccalaureate certificate	100	1.34	100	1.26
Master’s degree	1,910	19.85	1,930	19.38
Postmaster’s certificate	40	0.39	40	0.36
Professional degree	390	4.01	400	3.90
Doctor’s degree	210	2.14	210	2.00

See notes at end of table.

Table 40. Before and after imputation distributions of derived categorical variables: 2003—Continued

Variable	Before imputation		After imputation	
	Sample size	Weighted percent	Sample size	Weighted percent
Had ever enrolled in a degree program after BA in 1993 (B3ENRPG)	8,380	100.00	8,970	100.00
Graduate only	3,560	38.84	3,790	37.93
Undergraduate only	430	4.93	440	4.61
Both graduate and undergraduate	390	4.69	410	4.28
No enrollment	4,010	51.54	4,320	53.17
Currently enrolling in a degree program (B3CURENR)	8,930	100.00	8,970	100.00
Graduate only	700	7.09	710	7.11
Undergraduate only	80	0.96	90	0.97
Both graduate and undergraduate	#	0.03	10	0.04
No enrollment	8,140	91.92	8,170	91.88
Teacher pipeline status since graduation (B3PIPLIN)	8,880	100.00	8,970	100.00
No interest or action	3,880	47.59	3,910	47.33
Considered teaching previously	980	11.35	980	11.24
Considered teaching in 2003	1,220	13.57	1,220	13.45
Has applied	370	3.91	370	3.92
Taught, no training	340	3.43	350	3.45
Student, no teach, no certificate	210	2.05	220	2.31
Student, taught, no certificate	90	1.02	100	1.03
Certificate, no teach	190	2.00	210	2.26
Certificate, taught	1,610	15.07	1,620	15.01
Highest level of teacher certification (B3HICERT)	8,890	100.00	8,970	100.00
Never certified	7,010	81.98	7,030	81.56
Other	30	0.33	30	0.33
Emergency	30	0.35	30	0.35
Temporary	50	0.51	50	0.50
Probationary	50	0.51	50	0.50
Regular	1,350	12.84	1,400	13.24
Advanced	380	3.49	380	3.52
Teaching status as of 2003 interview (B3TCHST)	8,940	100.00	8,970	100.00
Currently teaching	1,100	10.54	1,100	10.55
Left teaching	950	8.93	960	8.94
Never taught	6,890	80.53	6,910	80.51
Ever completed student teaching (B3EVRSTD)	8,950	100.00	8,970	100.00
No	6,810	79.35	6,830	79.39
Yes	2,140	20.65	2,140	20.61
Control/sector of K-12 school most recently taught (B3MRSECT)	8,720	100.00	8,970	100.00
Public	1,350	13.00	1,550	14.49
Private	240	2.46	290	2.82

See notes at end of table.

Table 40. Before and after imputation distributions of derived categorical variables: 2003—Continued

Variable	Before imputation		After imputation	
	Sample size	Weighted percent	Sample size	Weighted percent
Control/sector of K-12 school most recently taught (B3MRSECT) —Continued				
Preschool or postsecondary	10	0.12	10	0.13
Medical or correctional	10	0.03	10	0.09
Foreign	20	0.16	20	0.21
Not applicable	6,910	82.47	6,910	80.53
Uncodable	190	1.77	190	1.73
Preparation to teach (B3PRPTCH)	8,910	100.00	8,970	100.00
Certified	1,800	17.05	1,830	17.26
Student taught, not certified	300	3.08	310	3.34
Neither student-taught nor certified	6,810	79.87	6,830	79.39
Locale of school most recently taught (B3MRSLOC)	8,710	100.00	8,970	100.00
Large central city	220	2.27	270	2.62
Midsize central city	260	2.53	300	2.81
Urban fringe or large city	420	4.17	490	4.71
Urban fringe or midsize city	150	1.41	170	1.61
Large town	20	0.13	20	0.15
Small town	170	1.57	210	1.89
Rural	350	3.25	380	3.53
Not applicable	6,950	82.89	6,950	80.96
Uncodable	190	1.77	190	1.73
Level of school most recently taught (B3MRSLEV)	8,670	100.00	8,970	100.00
Elementary	1,010	9.71	1,140	10.56
Secondary	500	5.12	620	6.14
Combined	160	1.47	210	1.92
Not applicable	6,950	83.27	6,950	80.96
Uncodable	50	0.43	50	0.42
Date of first teaching job, all teachers (year part of B3SPSTRT)	8,490	100.00	8,970	100.00
1992	60	0.57	80	0.72
1993	610	5.32	790	6.97
1994	290	2.77	380	3.80
1995	140	1.29	190	1.75
1996	110	1.22	150	1.71
1997	80	0.75	110	0.96
1998	50	0.55	60	0.67
1999	60	0.58	80	0.73
2000	30	0.30	40	0.44
2001	40	0.47	60	0.65
2002	20	0.22	30	0.27
2003	10	0.09	10	0.09
Not applicable	7,000	85.86	7,000	81.24

#Rounds to zero.
NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

Table 41. Before and after imputation distributions of derived continuous variables

Variable	Before imputation		After imputation	
	Sample size	Weighted mean	Sample size	Weighted mean
Current salary 2003, all respondents (B3CRSAL)	8,540	$55,119.40	8,810	$55,029.49
Average academic year base salary (B3TSALR)	1,690	$32,914.56	1,960	$32,771.19
Percent free/reduced price lunch recipients, school most recently taught (B3MRSFLE)	1,170	31.78	1,600	32.08
Percent minority enrollment, school most recently taught (B3MRSMPC)	1,480	36.28	1,830	36.30

SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

6.5.1 Imputation of Interview Variables

The B&B:93/03 interview variables were separated into two groups depending on the respondent base (or variable conditions). The first, unconditional group consisted of variables that applied to all respondents. The second, conditional group consisted of variables that applied to only a subset of respondents. Within the unconditional group, variables were sorted by percent missing and then imputed in order, from lowest percent missing to highest. Within the conditional group, the variables were first sorted by conditionality and percent missing, then imputed in the appropriate sequence. Since all computer-assisted telephone interviewing (CATI) variables had less than 10 percent missing, a constant set of predictor variables was used in a CHAID analysis to determine imputation classes for each imputation variable. The analysis used the following set of predictor variables: age, gender, race/ethnicity, U.S. citizenship, dependency status, prior respondent, receipt of federal aid, and institutional region, institutional type, institutional level. Table 42 lists the imputation classes for each of the 27 CATI variables. Some of these predictor variables were missing for a small percentage of cases and were imputed first with a weighted sequential hot deck imputation.

6.5.2 Derived Variable Imputation

Derived variables for B&B:93/03 were imputed sequentially in four batches, using a specific order determined by the variable conditions resulting from the longitudinal nature of this study. Imputing sequentially allowed these derived variables (or further derived variables resulting from them) to be used as class variables for imputing variables in subsequent batches. The process helped to ensure consistency across derived variables.

Most of the derived variables had several constraints defined by different combinations of data collected in prior rounds of the study. Therefore, a procedure for finding appropriate donor cases was developed before the imputation was performed. The procedure involved defining mutually exclusive groups or classes of respondents that met the constraints. The groups were used as the imputation classes for the weighted sequential hot deck imputation procedure. For the derived variables that did not have any constraints, a CHAID analysis was performed. The predictor variables included any prior imputed variables, including interview variables. Table 43 provides a list of the derived variables, in the order in which they were imputed, along with a list of the other variables that were used to define imputation classes. Due to the complex nature of the constraints, a table of the imputation classes of derived variables is not provided.

Table 42. Imputation classes for B&B:93/03 interview variables: 2003

Variable	Imputation classes
Marital status (B3MAR)	Dependents, non-Hispanic Black
	Dependents, non-Hispanic White and American Indian/Alaskan Native
	Dependents, Hispanic or Asian/Native Hawaiian/Other Pacific Islander
	Independent without dependents, 27 years old or younger
	Independent without dependents, 28 years old or older
	Independent with dependents, White non-Hispanic, Hispanic, Asian/Native Hawaiian/Other Pacific Islander
	Independent with dependents, Black non-Hispanic or American Indian/Alaska Native
Community service or volunteer past year (B3COMSRV)	Female dependents, 22 years old and younger, in regions—Great Lakes, Plains, Southeast, Southwest, Rocky Mountains, and Far West
	Female dependents, 22 years old and younger, in regions—New England, Mideast, Far West, and Outlying Areas
	Female dependents, 23 years old and older
	Female independents with dependents, Black non-Hispanic, Hispanic
	Female independents with dependents, White non-Hispanic, Asian/Native Hawaiian/Other Pacific Islander, American Indian/Alaska Native
	Female independents without dependents
	Male dependents
	Male independents with dependents
	Male independents without dependents
Volunteer: education-related (B3VLTPA)	Female dependents
	Female independents
	Male, race other than White non-Hispanic, who did not receive federal aid
	Male, race other than White non-Hispanic, who received federal aid
	Male, White non-Hispanic
Volunteer: other work with kids (B3VLTPB)	Dependents, 23 years old or older, White non-Hispanic, Hispanic, Asian/Native Hawaiian/Other Pacific Islander
	Dependents, Black non-Hispanic and American Indian/Alaska Native
	Dependents, Male, 22 years old or younger, White non-Hispanic, Hispanic, Asian/Native Hawaiian/Other Pacific Islander
	Dependents, Female, 22 years old or younger, White non-Hispanic, Hispanic, Asian/Native Hawaiian/Other Pacific Islander
	Independents with dependents, 27 years old or younger
	Independents with dependents, Female, 28 years old or older
	Independents with dependents, Male, 28 years old or older
	Independents without dependents, Female
	Independents without dependents, Male
Volunteer: fundraising (B3VLTPC)	Male
	Female

See notes at end of table.

Table 42. Imputation classes for B&B:93/03 interview variables: 2003—Continued

Variable	Imputation classes
Volunteer: help for homeless/ community (B3VLTPD)	Male
	Female
Volunteer: service to the church (B3VLTPE)	Dependents in regions—Great Lakes and Plains
	Dependents in regions—Mideast, Far West, and Outlying Areas
	Dependents in regions—Southeast and Southwest
	Dependents in the New England region
	Independent with dependents in regions—Great Lakes, Plains, and Southwest
	Independent with dependents in regions—New England, Mideast, Far West, and Outlying Areas
	Independent with dependents in regions—Southeast and Rocky Mountains
	Independents without dependents
Volunteer: frequency (B3VLFRQ)	Dependents in regions—New England, Great Lakes, Southeast, Southwest, and Far west
	Dependents in regions—Mideast, Plains, and Outlying Areas
	Dependents in the Rocky Mountain region
	Independent without dependents, Female
	Independent without dependents, Male
	Independent without dependents, in public institutions
	Independent without dependents, in private institutions
Registered to vote (B3VTREG)	Asian/Native Hawaiian/Other Pacific Islander in all regions except Outlying Areas
	Hispanic and American Indian/Alaska Native, in all regions except Outlying Areas
	Hispanic, Asian/Native Hawaiian/Other Pacific Islander, American Indian/Alaska Native, in the Outlying Areas
	White non-Hispanic and Black non-Hispanic, 21 years old or younger
	White non-Hispanic and Black non-Hispanic, 22–27 years old
	White non-Hispanic and Black non-Hispanic, 28 years old or older
Voted in the November 2002 National Election (B3VTNEL)	27 years old or younger, Black non-Hispanic
	27 years old or younger, Hispanic, Asian/Native Hawaiian/Other Pacific Islander, American Indian/Alaska Native
	27 years old or younger, White non-Hispanic
	28 years old or older, Hispanic and American Indian/Alaska Native
	28 years old or older, White non-Hispanic, Black non-Hispanic and Asian/Native Hawaiian/Other Pacific Islander

See notes at end of table.

Table 42. Imputation classes for B&B:93/03 interview variables: 2003—Continued

Variable	Imputation classes
Political activities past two years (B3POLIT)	21 years old or younger in a private, for-profit institution
	21 years old or younger in a private, not-for-profit institution
	21 years old or younger in a public institution, Female
	21 years old or younger in a public institution, Male
	22-27 years old
	28 years old or older, Hispanic and Asian/Native Hawaiian/Other Pacific Islander
	28 years old or older, White non-Hispanic, Black non-Hispanic and American Indian/Alaska Native
Make a telephone call (B3TELPN)	23 years old or younger, Not a respondent to BB94 or BB97
	23 years old or younger, Respondent to either BB94 or BB97
	24–27 years old
	28 years old or older and did not receive federal aid
	28 years old or older, did receive federal aid, and attend a private institution
	28 years old or older, did receive federal aid, and attend a public institution
Undergraduate value: particular major(s) chosen (B3UGVLA)	Attend a public institution
	Attend a private, not-for-profit institution and receive federal aid
	Attend a private, not-for-profit institution and do not receive federal aid
	Attend a private, for-profit institution, Female
	Attend a private, for-profit institution, Male
Undergraduate value: liberal arts courses taken (B3UGVLB)	Attend a private, for-profit institution
	Attend a private, not-for-profit institution, 21 years old or younger
	Attend a private, not-for-profit institution, 22 years old
	Attend a private, not-for-profit institution, 23–27 years old
	Attend a private, not-for-profit institution, 28 years old or older
	Attend a public institution, 21 years old or younger
	Attend a public institution, 22–27 years old
	Attend a public institution, 28 years old or older
Undergraduate value: professional courses taken (B3UGVLC)	Dependent, Attend a private institution, 22 years old or younger
	Dependent, Attend a private institution, 23 years old or older
	Dependent, Attend a public institution, 21 years old or younger
	Dependent, Attend a public institution, 22 years old or older
	Independent with dependents, Female, Black non-Hispanic and Hispanic
	Independent with dependents, Female, White non-Hispanic, Asian/Native Hawaiian/Other Pacific Islander, and American Indian/Alaskan Native
	Independent with dependents, Male
	Independent without dependents

See notes at end of table.

Table 42. Imputation classes for B&B:93/03 interview variables: 2003—Continued

Variable	Imputation classes
Undergraduate value: quality of instruction (B3UGVLD)	Attend a private institution
	Attend a public institution, in regions—New England, Mideast, Great Lakes, and Plains, Dependent
	Attend a public institution, in regions—New England, Mideast, Great Lakes, and Plains, Independent with dependents, 22 years old and younger
	Attend a public institution, in regions—New England, Mideast, Great Lakes, and Plains, Independent with dependents, 23 years old and older, Did not receive federal aid
	Attend a public institution, in regions—New England, Mideast, Great Lakes, and Plains, Independent with dependents, 23 years old and older, Received federal aid
	Attend a public institution, in regions—New England, Mideast, Great Lakes, and Plains, Independent without dependents
	Attend a public institution, in regions—Southeast, Rocky mountains, and Far west, 23 years old or younger
	Attend a public institution, in regions—Southeast, Rocky mountains, and Far west, 24 years old or older
	Attend a public institution, in regions—Southwest and Outlying areas
Undergraduate value: internship and other work (B3UGVLE)	Female, Black non-Hispanic, Hispanic, and Asian/Native Hawaiian/Other Pacific Islander
	Female, White non-Hispanic and American Indian/Alaska Native, 21 years old or younger
	Female, White non-Hispanic and American Indian/Alaska Native, 22–23 years old
	Female, White non-Hispanic and American Indian/Alaska Native, 24 years old or older, Did not receive federal aid
	Female, White non-Hispanic and American Indian/Alaska Native, 24 years old or older, Received federal aid
	Male, Black non-Hispanic, Asian/Native Hawaiian/Other Pacific Islander, and American Indian/Alaska Native
	Male, White non-Hispanic and Hispanic
Undergraduate value: none of the above (B3UGVLF)	Attend a private, for-profit institution
	Attend a private, not-for-profit institution, 22 years old or younger
	Attend a private, not-for-profit institution, 23 years old or older
	Attend a public institution
Undergraduate preparation: work and career (B3UGPRA)	Dependent
	Independent with dependents, Did not receive federal aid
	Independent with dependents, Received federal aid
	Independent without dependents

See notes at end of table.

Table 42. Imputation classes for B&B:93/03 interview variables: 2003—Continued

Variable	Imputation classes
Undergraduate preparation: further education (B3UGPRB)	21 years old or younger, attend a private, for-profit institution
	21 years old or younger, attend a private, not-for-profit institution
	21 years old or younger, attend a public institution
	22–27 years old, Black non-Hispanic, Hispanic, and Asian/Native Hawaiian/Other Pacific Islander
	22–27 years old, White non-Hispanic and American Indian/Alaska Native
	28 years old or older, Did not receive federal aid
	28 years old or older, Received federal aid
Undergraduate preparation: financial security (B3UGPRC)	22 years old or younger
	23 years old or older, in New England region
	23 years old or older, in regions—Mideast, Plains, Southeast, Rocky Mountains, and Far west, Did not receive federal aid
	23 years old or older, in regions—Mideast, Plains, Southeast, Rocky Mountains, and Far west, Received federal aid
	23 years old, in regions—Great Lakes, Southwest, and Outlying Areas,
	24 years old or older, in regions—Great Lakes, Southwest, and Outlying Areas,
Undergraduate preparation: none of the above (B3UGPRD)	Black non-Hispanic
	Race other than Black non-Hispanic
Undergraduate education worth cost (B3UGWRA)	Attended a private institution, 27 years old or younger
	Attended a private institution, 28 years old or older, Did not receive federal aid
	Attended a private institution, 28 years old or older, Received federal aid
	Attended a public institution, Dependent
	Attended a public institution, Independent with dependents
	Attended a public institution, Independent without dependents
Undergraduate education worth time (B3UGWRB)	21 years old or younger
	22 years old, Hispanic and Asian/Native Hawaiian/Other Pacific Islander
	22 years old, White non-Hispanic, Black non-Hispanic and American Indian/Alaska Native
	23–27 years old
	28 years old or older, Dependent and Independent with dependents
	28 years old or older, Independent without dependents
Undergraduate education worth effort (B3UGWRC)	Female, Dependent, White non-Hispanic and American Indian/Alaska Native
	Female, Independent
	Female, Dependent, Black non-Hispanic, Hispanic, and Asian/Native Hawaiian/Other Pacific Islander
	Male

See notes at end of table.

Table 42. Imputation classes for B&B:93/03 interview variables: 2003—Continued

Variable	Imputation classes
Current employment status (B3CUREMP)	Female, 27 years old or younger, Black non-Hispanic and American Indian/Alaska Native
	Female, 27 years old or younger, Hispanic and Asian/Native Hawaiian/Other Pacific Islander
	Female, 27 years old or younger, White non-Hispanic, in regions—Great Lakes, Plains, Southwest, and Outlying Areas
	Female, 27 years old or younger, White non-Hispanic, in regions—New England and Mideast
	Female, 27 years old or younger, White non-Hispanic, in regions—Southeast and Far West
	Female, 27 years old or younger, White non-Hispanic, in Rocky Mountain region
	Female, 28 years old or older, Did not receive federal aid
	Female, 28 years old or older, Received federal aid
	Male, 27 years old or younger
	Male, 28 years old or older
Income from work in 2002 (B3INC02)	Female, in New England region
	Female, in regions—Mideast, Great Lakes, Far west, and Outlying areas
	Female, in regions—Plains, Southeast, and Southwest
	Female, in Rocky Mountain region
	Male, Dependent
	Male, Independent with dependents
	Male, Independent without dependents

NOTE: The following sort variables were used for all imputations: age as of 12/31/92 (AGE), adjusted institution stratum (BNBSTRAT), type of institution and enrollment category (ENRLLCAT), state of current residence (B3STATE), major field of study (12 categories) (MAJORS3), highest degree received after BA completion (B2HDGPRG), and currently enrolled in graduate program (B3CRGRD1).

SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

Table 43. Variables used for imputation classes for derived imputation variables

Derived variable	Variables used to define imputation classes	
Batch 1		
Labor force participation 2003 (B3LFP03)	B3CUREMP	Current employment status
	B3CURHRS	Current job: hours per week
Current salary 2003, all respondents (B3CRSAL)	GENDER2	Gender
	DEPEND2	Dependency status for financial aid
	FEDAID	Receipt of federal aid 1=Yes 2=No
	OBEREG	Institution Region
	AGEGROUP	Age Group 1=<22 2=22 3=23 4=24-27 5=>27
Occupational category (collapsed) (B3OCCAT)	GENDER2	Gender
	DEPEND2	Dependency status for financial aid
	RACEETH	Race/Ethnicity
	CONTROL	Institution Control
	AGEGROUP	Age Group 1=<22 2=22 3=23 4=24-27 5=>27
	FEDAID	Receipt of federal aid 1=Yes 2=No
Total number of dependent children in 2003 (B3NUMCH)	B3CHCNUM	Number of dependents age 0-4 in daycare
	B3D2AG1	Dependents outside household 0-4
	B3D2AG2	Dependents outside household 5-17
	B3D3AG1	Number of dependents ages 0-4
	B3D3AG2	Number of dependents ages 5-17
	B3DPAG1	Number supported financially ages 0-4
	B3DPAG2	Number supported financially ages 5-17
Wrote letter or e-mail to public official 2003 (B3WROTE)	AGEGROUP	Age Group 1=<22 2=22 3=23 4=24-27 5=>27
	RACEETH	Race/Ethnicity
	CONTROL	Institution Control
Highest degree attained (B3HDG03)	B2ATTENR	Attainment and enrollment
	B2HDGPRG	Highest degree received after BA completion
	B3AGEDOC	Age at doctor's degree receipt
	B3AGEMA	Age at master's degree receipt
	B3AGEPRO	Age at first-professional degree receipt
	B3ATTDOC	Post-BA enrollment: doctor's degree
	B3ATTMA	Post-BA enrollment: master's degree
	B3ATTPRO	Post-BA enrollment: professional degree
	B3CMPDOC	Post-BA completion: doctor's degree
	B3CMPMA	Post-BA completion: master's degree
	B3CMPPRO	Post-BA completion: professional degree
	B3GR1EMY	Graduate: date earned degree 1
	B3GR2EMY	Graduate: date earned degree 2
	B3GR3EMY	Graduate: date earned degree 3
	B3GRER1	Graduate: already received degree 1
	B3GRER2	Graduate: already received degree 2
	B3GRER3	Graduate: already received degree 3
	B3HDGPG	Highest degree program enrolled in after BA
	B3NCHDG	Highest graduate enrollment: completion rate
	B3RECPG	Type of degree program for the most recent graduate completion

See notes at end of table.

Table 43. Variables used for imputation classes for derived imputation variables—Continued

Derived variable	Variables used to define imputation classes	
Batch 1—Continued		
Had ever enrolled in a degree program after BA in 1993 (B3ENRPG)	B2ATTENR	Attainment and enrollment
	B2GRDST	Earliest graduate school start date
	B2HDGPRG	Highest degree received after BA completion
	B2HDGTIM	Attendance, highest post-BA attainment
	B2HENTIM	Attendance, highest post-BA enrollment
	B2P01CPR	Program type currently enrolled-school 1
	B2P01FCP	Current first program type at school 1
	B2P01PRG	Program enrolled at post-BA school 1
	B2P01SCP	Current second program type at school 1
	B2P02CPR	Program type currently enrolled-school 2
	B2P02FCP	Current first program type at school 2
	B2P02PRG	Program enrolled at post-BA school 2
	B2P02SCP	Current second program type at school 2
	B2P03CPR	Program type currently enrolled-school 3
	B2P03FCP	Current first program type at school 3
	B2P03PRG	Program enrolled at post-BA school 3
	B2P03SCP	Current second program type at school 3
	B2P03SCP	Current second program type at school 3
	B2P04PRG	Program enrolled at post-BA school 4
	B2P05PRG	Program enrolled at post-BA school 5
	B2RCNPRG	Degree program for most recent post-BA e
	B2RCNTIM	Attendance, most recent post-BA enroll
	B3AGEDOC	Age at doctor's degree receipt
	B3AGEMA	Age at master's degree receipt
	B3AGEPRO	Age at first-professional degree receipt
	B3ATTDOC	Post-BA enrollment: doctor's degree
	B3ATTMA	Post-BA enrollment: master's degree
	B3ATTPRO	Post-BA enrollment: professional degree
	B3CMPDOC	Post-BA completion: doctorate degree
	B3CMPMA	Post-BA completion: master's degree
	B3CMPPRO	Post-BA completion: professional degree
	B3GRAD	Attended a formal graduate program
	B3HDG03	Highest degree attained by 2003
	B3HDGMAJ	Highest graduate completion: major field of study
	B3HDGPRG	Highest graduate completion: degree program type
	B3HENPRG	Highest graduate enrollment: degree program type
	B3RECPG	Type of degree program for the most recent graduate completion
	B3UG	Attended a formal undergraduate program
	PB01PROG	Program type enrolled at postbaccalaureate school (1)
	PB02PROG	Program type enrolled at postbaccalaureate school (2)
	PB03PROG	Program type enrolled at postbaccalaureate school (2)
	YRENROL	Month first enrolled in graduate school

See notes at end of table.

Table 43. Variables used for imputation classes for derived imputation variables—Continued

Derived variable	Variables used to define imputation classes	
Batch 1—Continued		
Currently enrolling in a degree program (B3CURENR)	B3ENRPG	Had ever enrolled in a degree program after BA in 1993
	B3CRGRD1	Currently enrolled in graduate program 1
	B3CRUG1	Undergraduate 1: currently enrolled
Teacher pipeline status since graduation (B3PIPLIN)	B2PIPLIN	School 5 second period enroll status
	B2REGTJB	Number of regular teaching positions
	B2SECTOR	Got research assistantship for graduate
	B2SIZECL	Total number of students taught
	B2SLEVEL	Level of schools at which taught by 1997
	B2SPEND	Teaching status at end of study period
	B2SPEXPY	Total teaching experience in years
	B2SPNUM	Number of teaching spells
	B2SPSTD	Entered teaching during 1994-97
	B2SPSTI	Entered teaching during 1992-94
	B2SPSTYR	Year first began teaching after BA
	B2TCHTRN	Respondent took student teaching
	B2TJAPPL	Num. of teaching applications
	B3APPLY	Applied for a teaching job
	B3CONSDR	Currently considering teaching
	B3CRTTYP	Highest teaching certificate held
	B3EVRCON	Ever considered teaching
	B3EVRCRT	Ever licensed/certified to teach
	B3INDUCT	First: formal induction program
	B3LEAVA	Leave: low pay
	B3LEAVB	Leave: sabbatical or break
	B3LEAVC	Leave: change jobs out of education
	B3LEAVD	Leave: change jobs within education
	B3LEAVE	Leave: raise a family
	B3LEAVF	Leave: low prestige
	B3LEAVG	Leave: difficulty with people
	B3LEAVH	Leave: lack of autonomy
	B3LEAVI	Leave: accountability test scores
	B3LEAVJ	Leave: other
	B3LEAVW	Leave: none of the above
	B3PIPLIN	Teacher pipeline status
	B3PRPSTD	Completed student teaching
	B3TCHPRP	Done anything to prepare self to teach
	TEACHTRN	Ever taught or considered teaching
Highest level of teacher certification (B3HICERT)	B3PIPLIN	Teacher pipeline status
	B3CURCRT	Currently licensed/certified to teach
	B3EVRCRT	Ever licensed/certified to teach

See notes at end of table.

Table 43. Variables used for imputation classes for derived imputation variables—Continued

Derived variable	Variables used to define imputation classes	
Batch 2		
Teaching status as of 2003 interview (B3TCHST)	B2REGTJB	Number of regular teaching positions
	B2SECTOR	Sector of most recent school
	B2SIZECL	Total number of students taught
	B2SLEVEL	Level of schools at which taught by 1997
	B2SPEND	Teaching status at end of study period
	B2SPEXPY	Total teaching experience in years
	B2SPNUM	Number of teaching spells
	B2SPSTD	Entered teaching during 1994-97
	B2SPSTI	Entered teaching during 1992-94
	B2SPSTYR	Year first began teaching after BA
	B3DTFSTJ	Date of first teaching job
	B3PIPLIN	Teacher pipeline status
	B3TSALR	Academic year base salary, most recent teaching job
	B3TCHPST	Ever worked as teacher or aide
Ever completed student teaching (B3EVRSTD)	B3PIPLIN	Teacher pipeline status
	B3PRPSTD	Completed student teaching
	B3CRTTYP	Highest teaching certificate held
	B2PIPLIN	School 5 second period enroll status
	B3HICERT	Highest level of certification ever held
	B2TCHTRN	Respondent took student teaching
Average academic year base salary (B3TSALR)	GENDER2	Gender
	Y_SCLEV1	School 1 level
	B3CURENR	Currently enrolling in a degree program
	B2HDGPRG	Highest deg received after BA completion
	B3ENRPG	Had ever enrolled in a degree program after BA in 1993
	B3HICERT	Highest level of certification ever held
Batch 3		
Control/sector of school most recently taught (B3MRSECT)	GENDER2	Gender
	CONTROL	Institution Control
	DEPEND2	Dependency status for financial aid
Preparation to teach (B3PRPTCH)	B3CRTTYP	Highest teaching certificate held
	B2HICERT	Highest teacher certification type in B9
	B3PIPLIN	Teacher pipeline status
	B3EVRSTD	Completed student teaching
	B2TCHTRN	Respondent took student teaching
Batch 4		
Locale of school most recently taught (B3MRSLOC)	B3EVRSTD	Completed student teaching
	B3TCHST	Teaching status as of 2003 interview
Level of school most recently taught (B3MRSLEV)	B3EVRSTD	Completed student teaching
	B3PIPLIN	Teacher pipeline status
	GENDER2	Gender

See notes at end of table.

Table 43. Variables used for imputation classes for derived imputation variables—Continued

Derived variable	Variables used to define imputation classes	
Batch 4—Continued		
Percent minority enrollment, school most recently taught (B3MRSMPC)	B3EVRSTD	Completed student teaching
	B3TCHST	Teaching status as of 2003 interview
	RACEETH	Race/Ethnicity
	B3TSALR	Academic year base salary, most recent teaching job
	B3MRSECT	Sector of most recent school
Percent free/reduced price lunch recipients, school most recently taught (B3MRSFLE)	B3EVRSTD	Completed student teaching
	B3TCHST	Teaching status as of 2003 interview
	RACEETH	Race/Ethnicity
	B3TSALR	Academic year base salary, most recent teaching job
	DEPEND2	Dependency status for financial aid
Date of first teaching job, all teachers (B3SPSTRT)	B2REGTJB	Number of regular teaching positions
	B2SECTOR	Got research assistantship for graduate
	B2SIZECL	Total number of students taught
	B2SLEVEL	Level of schools at which taught by 1997
	B2SPEND	Teaching status at end of study period
	B2SPEXPY	Total teaching experience in years
	B2SPNUM	Number of teaching spells
	B2SPSTD	Entered teaching during 1994-97
	B2SPSTI	Entered teaching during 1992-94
	B2SPSTYR	Year first began teaching after BA
	COMPDAT2	Date interview completed

SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

6.5.3 Definition of Missing for Item Imputation

For B&B:93/03, several codes were used to distinguish between legitimate and nonlegitimate missing items for the purpose of imputation. These codes were the following:

- **Refusal/terminated early:** For all questions where the respondent refused to answer, the value of missing was assigned initially, and then imputed. Likewise, if the interview was terminated early, leaving some questions not asked, the value of missing was initially assigned and then imputed.
- **Legitimate skip**: Many respondents could legitimately skip questions that did not apply to them. In these cases, the missing responses were coded as legitimate skips and were not imputed.
- **Not applicable:** For B&B:93/03 nonrespondents, questions were not imputed.

6.5.4 Evaluation of Imputations

Comparing imputation distributions within imputation classes is a key measure for determining whether or not the weighted sequential hot deck imputation procedure produced acceptable results. The more similar the distributions, the more successful the imputation

process. For evaluation of the B&B:93/03 imputation results, distributions were considered to be similar when absolute differences were less than 5 percent. Absolute difference was calculated by subtracting the before imputation weighted percent from the after imputation weighted percent. If absolute differences greater than 5 percent were found, then the unweighted distributions would be examined to see if the large differences were due to small sample sizes. Any large differences would be evaluated and corrected (by using different imputation classes) and documented when no resolution was possible.

Tables 39 lists the before and after imputation distributions of the CATI categorical variables that were imputed and Tables 40 and 41 list the before and after imputation distributions of derived variables that were imputed. No absolute differences greater than 5 percent were found for any comparison.

This page left intentionally blank.

References

Bradburn, E.M., Nevill, S., and Forrest Cataldi, E. (forthcoming). *Where Are They Now? A Description of 1992–93 Bachelor's Degree Recipients 10 Years Later* (NCES). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans*. Philadelphia: Society for Industrial and Applied Mathematics.

Folsom, R.E., and. Singh, A.C. (2000). The Generalized Exponential Model for Sampling Weight Calibration for Extreme Values, Nonresponse, and Poststratification. *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, (pp. 598-603). American Statistical Association. Available: *http://amstat.org/sections/srms/proceedings/paper/2000_099.pdf*

Francisco, C.A. and Fuller, W.A. (1991). Quantile Estimation with a Complex Survey Design. Annals of Statistics, 19:454–69.

Green, P.J., Meyers, S.L., Giese, P., Law, J., Speizer, H.M., and Tardino, V.S. (1996). *Baccalaureate and Beyond Longitudinal Study: 1993/94 First Follow-up Methodology Report* (NCES 96–149). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Green, P.J., Meyers, S.L., Veldman, C., and Pedlow, S. (1999). *Baccalaureate and Beyond Longitudinal Study: 1993/97 Second Follow-up Methodology Report* (NCES 1999–159). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Henke, R., and Bugarin, R. (forthcoming). *Teaching Status of 1992–93 College Graduates as of 2003* (NCES). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Loft, J., Riccobono, J.A., Whitmore, R., Fitzgerald, R., and Berkner, L.K. (1995). *Methodology Report for the National Postsecondary Student Aid Study, 1992–93* (NCES 95–211). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Research Triangle Institute. (2001). *SUDAAN User's Manual, Release 8.0.* Research Triangle Park, NC: Research Triangle.

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

Wine, J.S., Cominole, M.B., Carwile, S., Franklin, J., Carley-Baxter, L., and Wheeless, S.C. (2004). *Baccalaureate and Beyond Longitudinal Study—Third Follow-Up (B&B:93/03) Field Test Methodology Report* (NCES 2004-02). U.S. Department of Education, National Center for Education Statistics. Washington, DC: National Center for Education Statistics.

Wine, J.S., Heuer, R.E., Wheeless, S.C., Francis, T.L., Franklin, J.W., and Dudley, K.M. (2002). *Beginning Postsecondary Students Longitudinal Study 1996-2001 (BPS:1996/2001) Methodology Report* (NCES 2002-171). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Woodruff, R.S. (1971). A Simple Method for Approximating the Variance of a Complicated Estimate. *Journal of the American Statistical Association,* 66:411–414.

Wölter, K.M. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.