National Center for Education Statistics

U.S. Department of Education
Institute of Education Sciences
NCES 2004–02

Working Paper Series

1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03) Field Test Methodology Report

November 2004

The Working Paper Series was initiated to promote the sharing of the valuable work experience and knowledge reflected in these preliminary reports. These reports are viewed as works in progress, and have not undergone a rigorous review for consistency with NCES Statistical Standards prior to inclusion in the Working Paper Series.

U.S. Department of Education
Rod Paige
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Education Statistics
Robert Lerner
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, the Congress, the states, 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 2004

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., Cominole, M., Carwile, S., Franklin, J., Carley-Baxter, L., and Wheeless, S. (2004). *1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03) Field Test Methodology Report* (NCES 2004–02). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Content Contact:
Kristin Perry
B&B Project Officer
Postsecondary Longitudinal & Sample Survey Studies-PSD
(202)502-7428
Kristin.Perry@ed.gov

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 cross-sectional study of how postsecondary students and their families pay for education beyond high school. A year later, a follow-up interview was conducted (B&B:93/94) and transcripts were collected from students' undergraduate institutions and coded. 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 report describes the methodology and findings of the B&B:93/03 field test interview, conducted in the spring and early summer of 2002, with 1991–92 bachelor's degree recipients. Prior to 2002, this field test cohort was interviewed in 1992 for the NPSAS:93 field test, and again in 1993 and 1996 for the B&B:93/94 and B&B:93/97 field test studies, respectively. The purpose of the B&B:93/03 field test was to evaluate the operational and methodological procedures, instruments, and systems planned for the B&B:93/03 full-scale study to ensure their efficiency in achieving the desired response rates and level of data quality.

Study Design

The target population for the B&B:93 field test set of studies consisted of those students who were eligible to participate in NPSAS:93 and were awarded the bachelor's degree during the 1991–92 academic year by a postsecondary institution in the United States, the District of Columbia, or Puerto Rico. The B&B field test cohort consisted of students who participated in the NPSAS:93 field test and were identified as baccalaureate recipients. In addition, the cohort retained some NPSAS:93 field test nonrespondents who were potentially eligible for B&B. Their eligibility was determined as part of the field test interview for the 1-year follow-up, B&B:93/94. The sample for the B&B:93/03 field test consisted of all respondents to the field test interview for the second follow-up, B&B:93/97, plus a subsample of nonrespondents, for a final starting sample of 925 college graduates.

For the first time, the B&B:93 field test offered sample members three response choices—a self-administered web interview, a telephone interview, and an in-person interview—programmed as a single web-based interview for use in all three modes of interviewing. Early in data collection, only the self-administered interview option was available, supported by specially trained Help Desk staff who could answer questions and handle problems as they arose. An early response incentive experiment tested whether or not respondents could be encouraged to complete the self-administered web option before telephone interviewing began. Ten days after the self-administered interview was made available, telephone interviewing began with those who had not already completed the interview. About 5 weeks later, field interviewing was begun. Only nonrespondents whose most recent address fell within one of seven predetermined

geographic clusters could be interviewed in person. Those cases not within a cluster were sent for additional tracing and continued attempts at telephone interviewing. Sample members could complete the self-administered interview at any time during data collection.

Instrumentation

For the first time, students were offered the opportunity to participate in the 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, a web site was developed to launch the self-administered interview, to provide additional study information, and to collect updated student locating information. Prior to the start of data collection, several steps were taken to ensure the usability of the field test instrument, including usability testing, evaluating on-screen motivators, and developing effective on-screen help text and coding systems.

The interview focused on students' activities in the 6 years since the last follow-up interview, B&B:93/97. Questions focused on education pursued since the last interview; current employment, employment patterns, and career development, with specific questions focused on the employment patterns and job satisfaction of new, current, and former teachers; background characteristics, such as marital status, family, civic involvement, and disability status; and finances, including education loan debt, assets, and income.

Data Collection Design and Outcomes

Training

Field test training programs were developed for staff monitoring the Help Desk (which was made available to sample members completing the self-administered interview) and for interviewers conducting telephone and in-person interviews. Training topics included an overview of B&B:93/03, solutions to common problems encountered using the self-administered interview, case management, quality control, establishing 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. In addition, tracing specialists received abbreviated training specific to the needs of locating B&B:93/03 sample members.

Interviewing

The self-administered field test interview was made available to sample members beginning in April 2002. Telephone interviewing of those sample members who had not yet completed the self-administered interview began 10 days later. Two months after the start of telephone interviewing, field interviewers began tracing and interviewing nonrespondents whose last known address was in one of seven geographic clusters. Data collection was concluded in early July 2002 once all data collection systems had been thoroughly tested.

From the starting sample of 925 members, 8 were found to be deceased or unavailable for the data collection period. Of the remaining sample members, 81 percent were contacted before the end of the period. Of those, 90 percent were interviewed for an overall response rate of 73 percent. Among respondents, 36 percent completed the self-administered interview on the

Internet, 53 percent completed a telephone interview, and the remainder were interviewed in person.

Interview Burden

The length of the B&B:93/03 field test interview was calculated separately according to whether the interview was self-administered or interviewer-administered. Self-administered interviews averaged 41 minutes, of which 12 minutes were 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 36 minutes, with about 4 minutes of this time required to transmit data (in the telephone interviews only, because the field interview was saved on and run from a stand-alone, laptop computer). While the overall time to complete the interview (including transit time) was longer for self-administered respondents, the actual time spent answering questions was shorter for self-administered respondents.

Usability of the Instrument

Based on the results of the usability testing conducted prior to the start of data collection, the B&B:93/03 field test instrument was revised to reduce the complexity of specific response tasks and to clarify the nature of data requested in specific items. During data collection, help text for every screen of the B&B:93/03 instrument could be displayed to provide instructions on how to enter responses, clarification of the type of information requested, and definitions of words or phrases within an item. Help text usage rates were fairly low across self-administered and interviewer-administered interviews.

The 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 for the correctness of codes selected by self-administered 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.

Early Response Incentive Experiment

An experiment was conducted during field test data collection to determine if the offer of a $20 monetary incentive would encourage sample members to complete the self-administered interview in the first 10 days of data collection, prior to the start of telephone interviewing. The sample was randomly divided into two groups. The control group received a mailing with information on how to complete the self-administered interview on the web. The experimental group received the same mailing, together with the offer of the incentive for completing the interview within 10 days. A comparison of response rates showed that the response rate for the incentive group was higher than for the control group, suggesting that payment of an early response incentive did increase the likelihood of an early response.

Indeterminate Responses

Every item in the interview allowed two types of specific nonresponse: "don't know" and "decline to answer" options. Overall, self-administered respondents were twice as likely as interviewer-administered respondents to provide an indeterminate response. Indeterminate

response rates varied considerably depending on the section of the interview and the nature of the questions being asked.

Recommendations for the Full-scale Study

The final chapter of the report summarizes changes recommended for the sampling design, tracing and data collection plans, and instrument for the full-scale data collection scheduled to begin in February 2003. For example, the progress bar presented on the self-administered interview screens will display progress within a section, while a list of sections above the bar will display progress across sections. Within the interview itself, data elements will be revised and a new approach implemented in an effort to reduce item-level nonresponse.

Working Paper Foreword

In addition to official NCES publications, NCES staff and individuals commissioned by NCES produce preliminary research reports that include analyses of survey results, and presentations of technical, methodological, and statistical evaluation issues.

The ***Working Paper Series*** was initiated to promote the sharing of the valuable work experience and knowledge reflected in these preliminary reports. These reports are viewed as works in progress, and have not undergone a rigorous review for consistency with NCES Statistical Standards prior to inclusion in the Working Paper Series.

Copies of Working Papers can be downloaded as pdf files from the NCES Electronic Catalog (http://nces.ed.gov/pubsearch/).

Marilyn M. Seastrom
Chief Mathematical Statistician
Statistical Standards Program

Ralph Lee
Mathematical Statistician
Statistical Standards Program

This page is intentionally left blank.

Foreword

This report describes the methods and procedures used for the field test data collection effort of the 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03). Students selected for the field test cohort, who earned their bachelor's degrees during the 1991–92 academic year, were first interviewed in 1992 as part of the field test component of the 1993 National Postsecondary Student Aid Study (NPSAS:93). They were interviewed a year later, in 1993, as part of the B&B:93/94 field test and again, in 1996, for the B&B:93/97 field test. The results reported here are for the field test component of B&B:93/03, the fourth and final interview with the B&B cohort, 10 years following degree completion.

The most significant difference between the B&B:93/03 follow-up interview and its predecessors is the option for sample members to complete a self-administered interview via the Internet. A single, web-based interview was designed and programmed for use by sample members, and by both telephone and field interviewers. The interview focused on additional education undertaken since the last interview in 1996, as well as employment status and career patterns. A separate section addressed movement into and out of the teacher pipeline for any sample members who have taught, are currently teaching, or are considering teaching as a career.

Evaluation of the procedures used in the B&B field test has led to refinements that will benefit the full-scale study implementation. We hope that the information provided here and in the full-scale methodology report will be useful to a wide range of interested readers, and that the results reported in the forthcoming full-scale descriptive summary report will encourage others to use the B&B data. We welcome recommendations for improving the format, content, and approach, so that future methodology reports will be more informative and useful.

C. Dennis Carroll
Associate Commissioner
Postsecondary Studies Division

This page is intentionally left 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 B&B Project Officer, who reviewed the report at various stages and provided useful feedback. We also wish to thank Maria Rojtman, Office of Postsecondary Education, U.S. Department of Education; Brian Trzebiatowski, American Association of State Colleges & Universities; and John Wirt, NCES, who reviewed 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, Research Triangle Institute (RTI) 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 to Lynne Hawley, 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/or locating information.

This page is intentionally left blank.

Table of Contents

List of Tables

This page is intentionally left blank.

List of Figures

This page is intentionally left blank.

Chapter 1
Overview of B&B:93/03

This document describes the procedures and results of the field test implementation of the 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03). Research Triangle Institute (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 [PL 103-382].

This introductory chapter provides an overview of the background, purposes, schedule, and products of the B&B:93/03 study, and the unique purposes of the field test. The second chapter describes the design and methods used during the field test. Outcomes of field test data collection are presented in chapter 3. Evaluations of the quality of data collected are provided in chapter 4, and recommendations for changes in design for the full-scale study are presented in chapter 5. Materials used during the field test are provided as appendixes to the report and cited, where appropriate, in the text.

A. Background and Objectives of B&B

The Baccalaureate and Beyond (B&B) longitudinal study follows a cohort of bachelor's degree recipients over time. Potential sample members were first identified for the B&B:93 cohort as part of the 1992–93 National Postsecondary Student Aid Study (NPSAS:93), a nationally representative, cross-sectional study of postsecondary students from among nearly all postsecondary institutions in the United States and Puerto Rico. To be eligible for 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 field test sample was the starting sample for all of the follow-up field test data collections, including that for B&B:93/03.

The NPSAS:93 base-year interview collected information on background characteristics, enrollment, employment, and education financing, including financial aid from students, institutions, and parents. Students eligible for the bachelor's degree at the time of the NPSAS:93 interview were identified for participation in the B&B longitudinal component and asked additional questions about their plans for the future, particularly graduate education and any plans to pursue a K-12 teaching career.

The B&B:93 cohort was first followed up 1 year following degree completion (B&B:93/94). The interview covered a number of topics related to education since the bachelor's degree, job search and the transition into employment, job training, family formation, civic participation, and finances, including income and student loan and other debt. In addition, transcripts were collected from the schools from which B&B sample members earned their bachelor's degrees. As part of the transcript data collection, school-level information, such as 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 sample 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 post baccalaureate enrollment, including degrees sought, enrollment intensity and duration, finances, and degree attainment. Employment information and experiences were also collected, such as the number of jobs held since the last interview, occupations, salaries and benefits, and job satisfaction. Those in or newly identified as being in 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), 10 years following degree completion, will allow further study of the issues already addressed by the preceding follow-up studies. The 2003 interview covers topics related to continuing education, degree attainment, employment, career choice, family formation, and finances. It also contains a separate set of questions directed at new entrants to the teacher pipeline as well as those who have left teaching since the last interview.

B. Overview of the Field Test

The main purpose of the field test was to use and evaluate all operational and methodological procedures, instruments, and systems planned for use in the full-scale study. Using and testing methodologies in the field test that parallel the data collection procedures proposed for the full-scale study allow such procedures to be adjusted as necessary, prior to the start of full-scale data collection. In B&B:93/03, a self-administered web option was offered for the first time to the B&B:93 cohort. Consequently, a number of issues related to web access, usability, and data quality had to be evaluated as part of the field test data collection.

The B&B field test sample was first selected as part of the NPSAS:93 field test interview conducted in 1992. Eligible sample members had to have earned a bachelor's degree at some time during the 1991–92 school year. Follow-up interviews with the field test sample occurred in 1993, 1 year following bachelor's degree attainment; in 1996, 4 years later; and for the current study, in 2002, 10 years later.

Data collection for the B&B:93/03 field test began with the self-administered web option only. Help Desk staff assisted respondents with any questions or problems using the web instrument, and an early response incentive was paid for web completes within the first 10 days of data collection. Telephone interviewing began at the end of that 10-day period. Field interviewing with computer-assisted telephone interview (CATI) nonrespondents was attempted if the sample member was last located in one of seven geographic clusters identified for the field test. The operational schedule for the B&B:93/03 study is presented in table 1.

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

Activity	Start date	End date
Field test		
Sampling	12/01	12/01
Tracing	9/01	6/02
Web/self-administered interviewing	4/02	7/02
Telephone interviewing	4/02	7/02
Field data collection	6/02	7/02
Data files and documentation	3/02	9/02
Field test report	6/02	6/03
Full-scale study		
Sampling	12/02	12/02
Tracing	10/02	9/03
Web/self-administered interviewing	2/03	9/03
Telephone interviewing	3/03	9/03
Field data collection	4/03	9/03
Data files and documentation	2/03	11/03
Methodology report	8/03	6/04
Data Analysis System (DAS)	6/03	6/04
Descriptive survey report	7/03	8/04

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

C. Products of B&B:93/03

Full-scale data, along with data from prior studies, will be used to examine a wide range of education policy questions. Public release data files will be constructed from the full-scale data and distributed to a variety of federal and private organizations and researchers. A number of reports, publications, or other public information releases are anticipated for B&B:93/03:

- a bibliography of publications using data from the B&B:93 cohort;
- methodology reports (one each for the field test and 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 summary of significant findings with an essay on a policy-relevant topic, such as the career paths and labor market experiences of those who did not pursue graduate education but immediately entered the labor force after earning the bachelor's degree. An additional essay will focus on movement into and out of the teacher pipeline over the 10 years since the degree.

This page is intentionally left blank.

Chapter 2
Design and Method of the Field Test

A. Sampling Design

The field test cohort of the B&B:93 longitudinal study, first selected as part of the field test for NPSAS:93, consists of students identified as baccalaureate recipients during the 1991–92 academic year. Postsecondary students eligible for the NPSAS:93 field test were those students who were

- enrolled in *either* (a) a course for credit toward a degree or formal award; (b) a degree or formal award program of at least 3-months duration; 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.

In addition, if they received a baccalaureate degree from an institution eligible for the NPSAS:93 field test between July 1, 1991 and June 30, 1992, they were also eligible for participating in the NPSAS:93 field test as well as to be a member of the B&B:93 field test cohort.

Since 1992 when the field test interview for NPSAS:93 was conducted, the B&B:93 field test cohort has been interviewed two additional times—in 1993 as part of the 1-year follow-up of graduates, and in 1996 as part of the 4-year follow-up. These two previous follow-up interviews included 1,850 baccalaureate degree recipients selected from the NPSAS:93 field test sample. The B&B:93/97 field test identified six ineligible cases among these 1,850 sample members, including two sample members who had died since the 1991–92 field test base-year interview and four who were found to have not graduated during the 1991–92 academic year. Another 13 sample members were found to be deceased during advance tracing activities. The sampling frame for the B&B:93/03 field test, therefore, included both respondents and eligible nonrespondents to the B&B:93/97 interview—1,383 eligible sample members who responded in 1996 and 448 eligible sample members who did not respond at that time.

The B&B:93/03 field test sample was selected using simple stratified random sampling. Strata for the B&B:93/97 field test *respondents* were based on

- control of the institution attended in the base year (public or private);
- age of the student as of December 31, 1992 (23 or less and 24 or older);

- elementary or secondary education "teacher pipeline" status (taught or considered teaching and did not consider teaching);[1] and
- advance tracing outcome from the 2002 tracing activities (located student, did not locate student but located other contact, located neither the student nor another contact).

The sample strata for the B&B:93/97 field test *nonrespondents* were based on

- control of the institution attended in the base year (public or private), and
- advance tracing outcome from the 2002 tracing activities (located student, did not locate student but located other contact, located neither the student nor another contact).

Stratification by type of institution and age of student was used to ensure that the field test sample members represent a sufficient range of respondent characteristics. Further stratification by "teacher pipeline" status was included, because the B&B longitudinal study focuses in part on students who entered the teaching profession or were considering entering at some point since earning the bachelor's degree. Stratification by advance tracing outcome was used to increase the field test response rate by oversampling students who were most likely to be located.

From the strata defined above, 850 respondents and 75 nonrespondents were selected for a total sample of 925 students. Since B&B:93/03 is the fourth and final interview of the B&B:93 cohort, the sample size of 925 graduates is sufficient to accomplish the field test objectives. The allocation of the field test sample is presented in table 2.

About 92 percent of the B&B:93/03 field test sample were *respondents* during the B&B:93/97 field test. Among the B&B:93/97 field test respondents, about two-thirds of the sample (n=567) consisted of the group who had taught or considered teaching, and the remaining one-third consisted of those who had not considered teaching (n=283). About 70 percent of the field test sample consisted of located sample members, 25 percent from the group for whom a contact was located, and the remaining 5 percent consisted of those who could not be located prior to the start of data collection. Located students were slightly oversampled, and the sample was allocated proportionately to the institutional control and age categories.

Among the B&B:93/97 *nonrespondents*, about 70 percent (n=52) were located sample members, 25 percent (n=19) were sample members for whom a contact was located, and the remaining 5 percent (n=4) were those who could not be located initially. Within these categories, the sample was allocated proportionately to the control of the stratum in the base year.

[1]A student was classified as a member of the teacher pipeline if there was evidence from the B&B:93/94 or B&B:93/97 field test that the student had either taught or considered teaching {if NTJOBS>0 or TEACHING=1 or BFORCERT=1 or TEACHEV=1 or TEACH=1 or TCHCONEV=1}. All others were classified as "didn't consider teaching."

Table 2. Sample allocation for the B&B:2003 field test

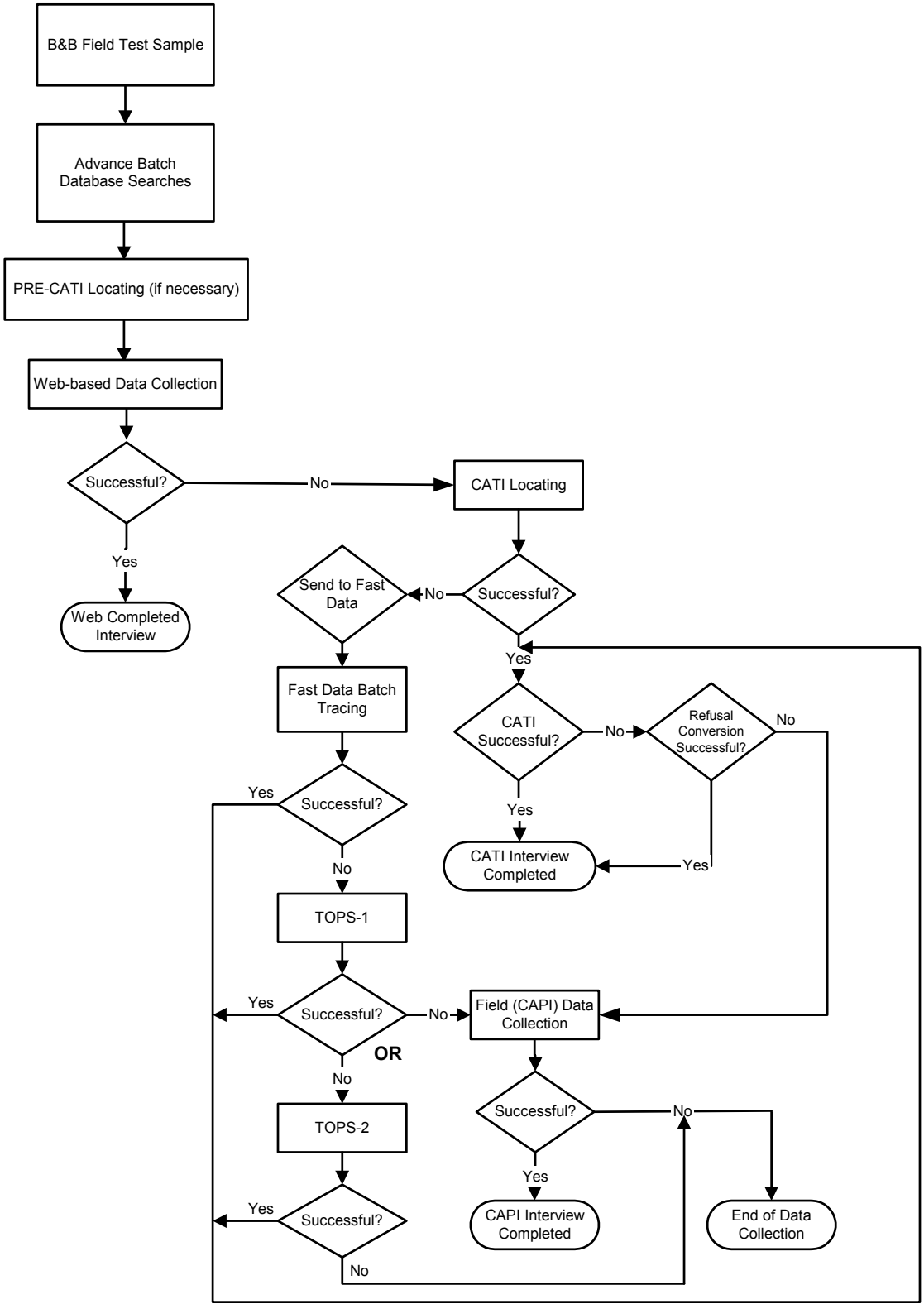
Stratum	B&B:93/97 FT response status	Teacher pipeline status	Advance tracing outcome	Control of base-year institution	Age in 1992	Frame count	Sample size
Total						1,831	925
1	Respondent	In teacher pipeline	Student located	Public	23 and younger	107	94
2	Respondent	In teacher pipeline	Student located	Public	24 and older	154	135
3	Respondent	In teacher pipeline	Student located	Private	23 and younger	138	121
4	Respondent	In teacher pipeline	Student located	Private	24 and older	54	47
5	Respondent	In teacher pipeline	Contact located	Public	23 and younger	67	41
6	Respondent	In teacher pipeline	Contact located	Public	24 and older	62	38
7	Respondent	In teacher pipeline	Contact located	Private	23 and younger	81	49
8	Respondent	In teacher pipeline	Contact located	Private	24 and older	23	14
9	Respondent	In teacher pipeline	Not located	Public	All ages	29	15
10	Respondent	In teacher pipeline	Not located	Private	All ages	25	13
	Respondent	Total In teacher pipeline				740	567
11	Respondent	Not in teacher pipeline	Student located	Public	23 and younger	94	47
12	Respondent	Not in teacher pipeline	Student located	Public	24 and older	127	64
13	Respondent	Not in teacher pipeline	Student located	Private	23 and younger	129	64
14	Respondent	Not in teacher pipeline	Student located	Private	24 and older	45	23
15	Respondent	Not in teacher pipeline	Contact located	Public	23 and younger	46	18
16	Respondent	Not in teacher pipeline	Contact located	Public	24 and older	49	19
17	Respondent	Not in teacher pipeline	Contact located	Private	23 and younger	77	30
18	Respondent	Not in teacher pipeline	Contact located	Private	24 and older	10	4
18	Respondent	Not in teacher pipeline	Not located	Public	All ages	27	6
20	Respondent	Not in teacher pipeline	Not located	Private	All ages	39	8
	Respondent	Total not in teacher pipeline				643	283
Total Respondents						1,383	850
21	Nonrespondent		Student located	Public	All ages	108	28
22	Nonrespondent		Student located	Private	All ages	91	24
23	Nonrespondent		Contact located	Public	All ages	71	11
24	Nonrespondent		Contact located	Private	All ages	55	8
25	Nonrespondent		Not located	Public	All ages	62	2
26	Nonrespondent		Not located	Private	All ages	61	2
Total Nonrespondents						448	75

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

B. Field Test Design

The activities of the B&B:93/03 field test were designed to fully test all procedures, methods, and systems of data collection prior to full-scale data collection. The remainder of this chapter provides an overview of the field test, organized around three main areas: pre-data collection activities, data collection activities, and the supporting data collection systems. The field test data collection process is depicted in figure 1.

Figure 1. Field test data collection process



NOTE: CATI=Computer-assisted telephone interview. CAPI=Computer-assisted personal interview. TOPS=Tracing Operations. Cases were sent for routine tracing (TOPS-1) if not located during initial telephone interviewing. Cases still not located, but not eligible for field data collection, were sent for intensive tracing (TOPS-2).
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

1. Pre-data Collection Activities

a. Advance tracing

Advance tracing activities for the B&B:93/03 field test were conducted prior to the start of data collection, so that new or updated locating information could be obtained for field test sample members. Sample member address files prepared by RTI programmers were sent for batch tracing using the Department of Education's Central Processing System (CPS) and the National Student Loan Data System (NSLDS), the National Change of Address (NCOA), TransUnion's credit information, Telematch, and ComServ's Death Information System (DIS) databases. For many sample members, these searches yielded new information or confirmed the original locating data. If batch locating efforts were unsuccessful, cases were sent to RTI's Tracing Operations (TOPS) unit for more advanced tracing.

b. Student web site

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

The B&B web site 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 web site. At the first tier, sample members were required to log on to the secure areas of the web site using a unique and randomly assigned study identification code 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 web site—both contact information and interview responses—were protected with Secure Socket Layer (SSL) technology, ensuring that only encrypted data were transmitted over the Internet.

c. Interview design

A single, web-based instrument was designed and programmed for the B&B:93/03 field test for use in three modes of data collection: as a self-administered interview via the Internet, as a computer-assisted telephone interview (CATI), and as a computer-assisted personal interview (CAPI). B&B sample members could access the interview directly from the study web site by entering the identification code and password provided in a mailing. Telephone interviewers accessed the web interview through the case management system in RTI's Telephone and Internet Operations (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.

The usability of the web instrument for self-administered interviewing was tested by RTI's Cognitive Laboratory with subjects in the same age range as the B&B:93 field test cohort and with a minimum education level of a bachelor's degree. Each subject was asked to complete the web interview while "thinking aloud" about such aspects of the interview as its appearance and ease of use. All subjects who participated used the same computer, operating system, Internet provider, and connection speed. The results of the usability testing, summarized later in this report, were provided to the instrument designers during development and incorporated into the design of the final instrument prior to the start of data collection.

d. Student mailings

Several weeks before the start of data collection, a prenotification mailing was sent to all sample members to provide information about the study and the start of data collection (samples of the prenotification mailing, and other student mail out materials, are provided in appendix A). This first mailing allowed RTI to evaluate the accuracy of the contact information obtained from advance tracing. The mailing included a letter, study leaflet, and an address update sheet with a postage-paid envelope. All materials provided sample members with the student web site address and a toll-free number to the B&B:93/03 study hot line. If the prenotification mailing was returned as "undeliverable," forwarding address information from the post office, if available, was entered into a locator database and used as the next best address for remailing.

One week before the start of data collection, a second packet was mailed to sample members that informed them of the start of data collection. The mailing included a personalized letter, the study web site address, a study leaflet, an address update sheet and postage-paid envelope, and a trinket. The trinket was a magnetic picture frame with an insert displaying the sample member's user name and password and the project's toll-free telephone number.

e. Early incentive experiment

As part of field test data collection, an experiment was conducted to determine whether payment of an incentive for early completion of the interview would have a positive effect on response rates and data collection costs. If sample members completed the B&B interview within the first 10 days of data collection, they received a $20 check. Only web self-administered interviewing was possible during the 10-day period. If a respondent called in with computer difficulty, however, the incentive was paid for completion of a telephone interview with Help Desk staff trained to conduct the web interview by telephone.

Prior to the start of data collection, the field test sample was divided into two incentive groups. Forty percent of the sample (370 cases) was assigned to a control group, which received a lead letter that asked them to complete the self-administered

interview with no offer of a monetary incentive. The remaining 60 percent of the sample (555 cases) received a similar lead letter with an offer to receive $20 if they completed the interview by April 25, 10 days following the start of data collection. Results of this early incentive experiment are discussed in chapter 3.

f. Staff training

Several different staff were trained to participate in the B&B data collection. Separate training sessions were conducted for tracing specialists, supervisors and monitors, Help Desk operators, telephone interviewers, and field interviewers. Training agendas are provided in appendix B. Common to each training session was a study overview, a review of the confidentiality requirements, a demonstration interview, a question-by-question review of the survey instrument, and hands-on practice exercises with the instrument, tracing module, and open-ended coding modules. Each training session was further specialized as follows:

- *Tracing specialists* received instruction on the most effective tracing sources as well as project-specific protocol on how to trace the sample members throughout the data collection period.
- *Help Desk operators* reviewed the "frequently asked questions," including responses to instrument-specific questions as well as technical issues, and instructions on how to document each call to the study hot line.
- *Telephone interviewers* were trained in techniques for gaining the cooperation of sample members, parents, and other contacts, as well as techniques for addressing the concerns of reluctant participants and refusal avoidance.
- *Field interviewers* were trained on field-specific operations, including the field case management system and field tracing procedures.

2. Data Collection Activities

a. Help Desk

A Help Desk was available to assist sample members who had questions or problems accessing and completing the self-administered interview. A toll-free hotline was set up to accept incoming Help Desk calls. If technical difficulties prevented a sample member from completing a self-administered interview, a Help Desk staff member, also trained to conduct telephone interviews, would encourage him/her to complete a telephone interview rather than to attempt the web interview.

A Help Desk application documented all calls from sample members. In addition, it provided

- information needed to verify a sample member's identity;
- login information allowing a sample member to access the web interview;
- systematic documentation of each call; and
- a method for tracking calls that could not be immediately resolved.

The Help Desk application also provided project staff with various reports on the types and frequency of problems experienced by sample members as well as a way to monitor the resolution status of all Help Desk inquiries.

b. Interviews

For the first time, B&B:93 field test sample members had the option of completing a self-administered interview via the Internet. The web interviewing option was introduced in the data collection notification mailing. For the first 10 days of data collection, only web interviews were completed unless a respondent called the Help Desk because of difficulty completing the interview over the Internet.

At the end of the 10-day web interviewing period, CATI interviewing began. The interviewer-administered interview was identical to the self-administered interview except that instructions to interviewers on how to administer each question were embedded at the top of each CATI screen in place of the progress bar. An automated call-scheduler, embedded within the CATI software, assigned cases to interviewers. This system allowed calls to be scheduled on the basis of case priority and time of day. Scheduler case assignment maximized the likelihood of contacting and interviewing sample members. If a web interview was in progress or had recently been completed, the call scheduler would prevent a CATI call to the same case. If a respondent told an interviewer that he/she preferred to complete the self-administered interview, interviewers would set a call back appointment for 2 weeks from the date of the original contact. If the self-administered interview had not yet been completed, the interviewer would attempt to complete a telephone interview.

When a sample member could not be located at a known address during CATI, interviewers conducted limited tracing using Fast Search and Transfer technology (FAST Data) and directory assistance services and by networking with other contacts provided by the sample member during an earlier interview. Cases that could not be located using any of the existing address information were identified for intensive tracing in RTI's Tracing Operations unit (referred to as TOPS1 in chapter 3). Cases that failed to be located a second time were either sent to the field for locating and interviewing, or returned to Tracing Operations for additional intensive tracing (TOPS2).

Field (CAPI) interviewing with nonrespondents began approximately 5 weeks following the start of CATI interviewing. Field interviews were conducted either in person or by telephone by the local field interviewer assigned to any one of seven geographic clusters: Washington, DC; New York, NY; Albion, RI; Detroit, MI; Flint, MI; New Orleans, LA; and Greensboro, NC. Cases assigned to the field could not be accessed by CATI interviewers but could still be completed as a self-administered interview over the Internet. Like the CATI interview, the CAPI interview presented interviewer instructions at the top of each screen.

For the field test, all nonrespondent cases thought to reside in one of the seven clusters were assigned to the field for interviewing. Cases were assigned based on the last known address for the sample member. Daily reports monitored each field interviewer's progress.

c. Nonresponse incentive

Use of incentives for nonrespondents has been shown to be effective in increasing response rates while containing data collection costs (Duffer et al. 1994). 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" (i.e., those with 15 or more call attempts and with whom contact had been established but no appointment scheduled). All cases assigned to field interviewers 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. The results of the early response incentive experiment are presented in section D of chapter 3.

C. Data Collection Systems

1. Instrument Design and Documentation System (IDADS)

The Instrument Design and Documentation System (IDADS) is a controlled web 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 Windows and web interfaces. There are three modules within IDADS: specifications, programming, and documentation.

Initial specifications were generated within the IDADS *specification module* that enabled access for searching, reviewing, commenting on, updating, exporting, and importing information associated with instrument development.

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 screen 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 program response-based routing.

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. Integrated Management System (IMS)

All aspects of the field test were under the control of an Integrated Management System (IMS). The IMS is 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 consists of several components, or modules: the management module, the Receipt Control System (RCS) module, and the CATI/CAPI module.

The *management* module of the IMS contains tools and strategies to assist project staff and the NCES project officer in managing the study. All information pertinent to the study is located there, accessible via the web, in a secure desktop environment. Available on the IMS are the current project schedule, monthly progress reports, daily data collection reports and status reports (available through the Receipt Control System 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 has a download area from which the client, contractors, and subcontractors can retrieve large files when necessary.

The RCS is an integrated set of systems that monitors all activities related to data collection, including tracing and locating. Through the RCS, project staff are able to perform stage-specific activities, track case statuses, identify problems early, and implement solutions effectively. The RCS's locator data are used for a number of daily tasks related to sample maintenance. Specifically, the mailout program produces mailings to sample members, the query system enables administrators to review the locator information and status for a particular case, and the mail return system enables project staff to update the locator database as mailings or reply sheets are returned or forwarding information is received.

A subcomponent of the RCS, the Field Case Management System (FCMS), controls field interviewing activities. The FCMS allows field staff to conduct tracing and interviewing activities, communicate with RTI staff via electronic mail, transmit completed cases, and receive new cases. The RCS also interacts with the Tracing Operations (TOPS) database, sending locator data between the two systems as necessary.

The CATI/CAPI module manages development of the CATI/CAPI instrument within IDADS. Developing the CATI/CAPI instrument with IDADS ensures that all variables are linked to their item/screen wording and thoroughly documented.

Chapter 3
Data Collection Outcomes

The data collection effort for the B&B:93/03 field test involved several steps, including attempting to locate sample members, initiating intensive locating efforts for hard-to-locate sample members, evaluating the utility of incentives paid for early response, and completing either a self-administered, telephone, or in-person interview. This chapter reports the outcomes of the field test, including interview rates overall and by mode, as well as burden on respondents. It also evaluates the effectiveness of the data collection methods employed in locating, contacting, incentivizing, and interviewing sample members.

A. Contacting and Interviewing Outcomes

Overall contacting and interviewing results for the B&B:93/03 field test are presented in figure 2. Of the 925 cases in the original sample, 748 were contacted (81 percent).[2] Eight of the sample members (0.9 percent) were excluded because they were deceased, out of the country, institutionalized, or physically or mentally incapacitated.[3] Of the 748 sample members contacted, 75 (10 percent) were nonrespondents at the end of data collection. About half of these nonrespondents refused to participate in the interview; for the other half, time ran out before an interview could be completed. For virtually all of the noncontact cases (97 percent), tracing was still underway when data collection ended.

Among the sample members contacted, 673 (90 percent) were interviewed. Full interviews were completed with 653 sample members (97 percent). The remaining 20 interviews were completed at least through the end of the first section of the interview (on education and training since the last interview) and, therefore, were considered partial interviews. The unweighted contact rate, not including exclusions, was 81.6 percent (748/917). For those contacted, the interview rate was 89.9 percent (673/748). The overall unweighted response rate was 73.4 percent (673/917).

1. Interviewing Outcomes by Mode

The B&B:93/03 field test used a web-based, multimode data collection strategy combining self-administered, telephone, and in-person interviewing options. At the start of data collection, only the self-administered web option was available. Help Desk staff were available both to assist with computer problems and to complete a telephone interview if the computer problems were insurmountable. Ten days following the start of self-administered interviewing, CATI began with all incomplete cases. About 1 month following the start of CATI, selected cases were sent to the field for in-person interviewing (CAPI).

---

[2] Defined as reaching the household of the sample member.
[3] One sample member was identified as physically or mentally incapacitated by parents or other contacts.

Figure 2. B&B:93/03 contacting and interviewing outcomes

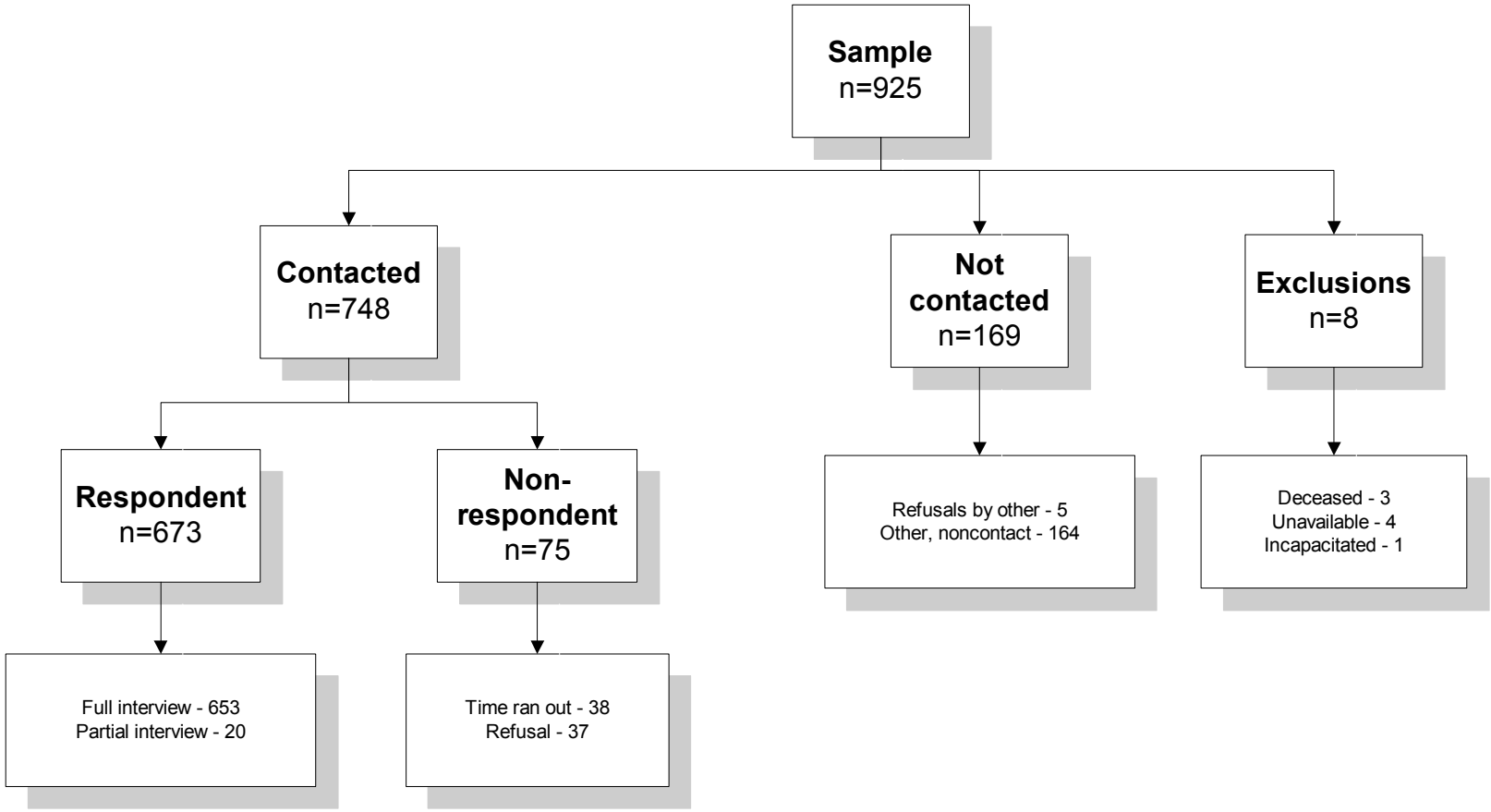


Table 3 presents the distribution of response status by mode of interview administration. Although a new option for sample members, 36 percent of interviews were completed using the web self-administered interview. Slightly more than half of all respondents (53 percent) completed a telephone interview, while only 12 percent of interviews were completed in the field.

Table 3. B&B:93/03 response status, by mode of administration

Mode of administration	Total sample	Percent of sample	Total respondents	Percent of all respondents
Total	917	100.0	673	100.0
Web	240	26.2	240	35.7
Computer-assisted telephone interview	353	38.5	353	52.5
Computer-assisted personal interview	80	8.7	80	11.9
Nonrespondents	244	26.6	†	†

†Not applicable.
NOTE: Sample members who are deceased, unavailable, or incapacitated 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).

The B&B:93/03 web instrument was made available on the study web site for self-administered interviewing beginning April 15, 2002. A total of 240 respondents (26 percent of the sample) completed the self-administered web interview, and web completes accounted for nearly 36 percent of all completed interviews. Fifteen of the interviews begun as web self-administered interviews were ultimately completed in CATI; 102 self-administered interviewers were completed within the 10-day, early response incentive period. Of those, 70 were incentivized cases (see section D below for a discussion of the incentive experiment results).

The Help Desk was opened at the same time that the web interview was made available. Nearly 20 percent of respondents who completed the interview on the web did so without calls to the Help Desk, interviewer prompting, or incentives. For the remaining 80 percent of web completes, one or more "interventions" (e.g., Help Desk assistance, interviewer prompting) was necessary to facilitate interview completion.

CATI locating and interviewing began on April 25, 2002, and continued for approximately 10 weeks. By the end of data collection, 353 interviews had been completed by telephone, including 5 partial interviews.

The last phase of field test data collection was CAPI. For the full-scale study, cases will typically be sent for CAPI when other tracing and interviewing efforts prove unproductive because the sample member (1) is unlocatable, (2) has been located but is unreachable by telephone, or (3) either explicitly refuses or expresses a preference for an in-person interview. In the field test, only 15 cases that met one of these three criteria were sent for CAPI, and 7 of those (47 percent) completed an interview.

In order to thoroughly test CAPI file transfer and data collection procedures prior to full-scale administration, an additional 120 pending cases were sent to the field if the sample member last resided within one of the seven geographic clusters identified for the field test. Cases within clusters were sent for CAPI interviewing irrespective of their status at the time of the assignment.

Of those cases, 73 (61 percent) completed the interview. The overall CAPI interview rate was 59 percent for the field test.

2. Interviewing Outcomes by Prior Response Status

For longitudinal studies, response status to a preceding interview is typically a good predictor of a sample member's likelihood to participate in the current interview. Table 4 shows the response status of B&B:93/03 field test sample members by their response status to the last follow-up interview, B&B:93/97.[4] Almost 76 percent of B&B:93/97 respondents participated in the B&B:93/03 interview. In contrast, only 48 percent of B&B:93/97 nonrespondents participated (χ^2=27.0, p<0.0001). Thus, respondents in B&B:93/97 were considerably more likely than nonrespondents to participate in B&B:93/03.

Table 4. B&B:93/03 response status, by prior response status

			Response status in B&B:93/03			
	Number sampled		Respondent		Nonrespondent	
Response status in B&B:93/97	Number	Percent	Number	Percent	Number	Percent
Total	917	100.0	673	73.4	244	26.6
Respondent	842	91.8	637	75.7	205	24.3
Nonrespondent	75	8.2	36	48.0	39	52.0

NOTE: Sample members who are deceased, unavailable, or incapacitated at the time of data collection 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).

B. Locating and Interviewing Outcomes

Effective tracing and locating of the 925 sample members was critical to the success of the B&B:93/03 field test data collection effort. Since the last contact with field test sample members occurred in 1996, or in 1993 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.

1. Pre-data Collection Tracing

For the field test, tracing began in the fall of 2001 by updating address and other contact information collected during the B&B:93/94 and B&B:93/97 field test interviews. Several tracing resources were used, including the Central Processing System (CPS), which contains federal financial aid application information, TransUnion's credit information, and databases from Telematch, DIS, and NCOA. All 1,850 potential field test cases were sent for batch tracing, and the sample was subsequently stratified and subsampled based, in part, upon the information gathered during batch tracing. Table 5 shows the record match rate for each method of batch tracing employed.

[4] Data are not available for this field test cohort concerning response status to the base-year study in 1993 and first follow-up interview in 1994.

Table 5. B&B:93/03 batch processing record match rates, by tracing source

Method of tracing	Number of records matched	Percent matched[1]
Central Processing System (CPS)	34	1.8
National Change of Address	308	16.6
Telematch	690	37.3
TransUnion	1,844	99.6
Death Information System	8	0.4

[1]Percent match rate is based on the 1,850 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.
NOTE: The final sample of 925 was selected using a stratified random sampling design from the 1,850 cases sent for pre-data collection batch tracing. 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).

Pre-data collection tracing continued with a mailing of prenotification materials to the 925 selected 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 complete an address update form either on the study web site or on a hardcopy form. Table 6 shows the contact and interview rates for those who returned some form of address update sheet. Almost all sample members who returned a hardcopy reply were interviewed (98.5 percent), and all of those who updated their contact information on the web completed the interview. Receipt of self-reported address updates was a good predictor of contacting and interviewing the sample member.

Table 6. B&B:93/03 contact and interview rates, by address update reply

Type of address update	Total	Percent contacted[1]	Percent interviewed, given contact[1]
Total	917	80.8	90.0
Hardcopy reply	69	97.1	98.5
Web site reply	33	100.0	100.0
None	815	78.7	88.6

[1] Percentages are based on the total within the row under consideration.
NOTE: Sample members who are deceased, unavailable, or incapacitated 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).

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 261 cases were sent to FAST Data, 75 percent of which were returned with new information or a verification of existing information. Of those cases, 66 percent were contacted, and 87 percent of those contacted were interviewed (57 percent overall interview rate).

3. Intensive Tracing

If a case could not be located through FAST Data, it was sent for intensive tracing conducted by RTI's Tracing Operations (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 field test 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 7 provides the results of the Tracing Operations intensive tracing efforts. Each phase of Tracing Operations tracing allowed for more in-depth tracing efforts. The most comprehensive stage of locating activities was used only when all previous tracing efforts failed and the sample member was not located in one of the seven geographic clusters identified for CAPI interviewing. As shown in table 7, of the cases traced, 66 percent were contacted, and of those, 85 percent were interviewed. This shows that once the member was located, the likelihood of completing an interview was high.

Table 7. B&B:93/03 contact and interview rates, by intensive tracing status

Intensive tracing status	Total	Contacted[1]		Interviewed, given contact[1]	
		Number	Percent	Number	Percent
Total	925	748	80.9	673	90.0
Cases requiring intensive tracing	282	187	66.3	159	85.0
Cases not requiring intensive tracing	643	561	87.2	514	91.6

[1] Percentages are based on the total within the 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).

C. Refusal Conversion Efforts

Refusal conversion procedures were used to gain cooperation from individuals who refused to participate in the field test 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 interview. The training emphasized how to gain cooperation, overcome objections, address the concerns of gatekeepers, and encourage participation.

Of the 748 sample members contacted, 129 cases (17 percent) were referred to refusal conversion specialists. Among the refusals, 51 cases (40 percent) were successfully converted, that is, the interviews were completed. Most of the converted interviews (60 percent) were completed by telephone.

Table 8 displays a breakdown of refusal conversion rates by prior response status. Conversion rates were higher among B&B:93/97 respondents than among nonrespondents. That

is, of the 112 B&B:93/97 respondents who refused to participate in the current (B&B:93/03) interview, 50 cases (45 percent) eventually completed. Of the 17 B&B:93/97 nonrespondents who refused, only 1 case (6 percent) completed the interview. Participation in prior B&B interviews tended to decrease the likelihood of refusing to be interviewed for B&B:93/03 (χ^2=5.17, p<0.05).

Table 8. B&B:93/03 refusal and refusal conversion rates, by prior response status

Prior response status	Total	Ever refused B&B:93/03 interview		Interviewed, after refusal	
		Number	Percent	Number	Percent
Total	925	129	13.9	51	39.5
1997 Respondents	850	112	13.2	50	44.6
1997 Nonrespondents	75	17	22.7	1	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).

D. Early Response Incentive Experiment

The B&B:93/03 field test included an experiment to determine if the offer of an incentive (i.e., a personalized check for $20) would increase the rate of early response to the self-administered web interview, ultimately decreasing data collection costs. Prior to the start of data collection, the field test sample was randomly divided into two groups. The control group (n=367) received all mailings described in chapter 2 and was informed of the web option but not offered the incentive. The incentive group (n=550) received the same mailings and information about the web option as did the control group. However, in the letter announcing the start of data collection, the incentive group was also offered $20 if the web self-administered interview was completed by April 25, 2002 (see appendix A for the data collection letters for each group).

A total of 102 web interviews (32 by the control group and 70 by the incentive group) were completed during the 10-day period. A comparison of response rates shows that the response rate for the incentive group (12.7 percent) was higher than for the control group (8.7 percent), suggesting that payment of an early response incentive does increase the likelihood of an early response (z=1.9; p<0.05).[5]

[5]The effect size is calculated using the difference between the arcsine transformation of the means of the incentive and control groups,€ divided by the standard deviation, or $\frac{\arcsin(\sqrt{p_i}) - \arcsin(\sqrt{p_c})}{1/2}$ (Cohen 1988). Using p_i=0.127 and p_c=.087 as the estimates for the incentive and control groups, respectively, the observed effect size is calculated to be 2×(.36438 - .29941) = 0.12994 or 0.13. By using the arcsine transformation, the effect of small proportions is removed from the variance and also increases the power. The power for detecting the difference at the ∀=0.05 level would be 0.61 using the transformation compared to 0.52 with, when using only the normal approximation to the proportions.

E. Nonresponse Incentive

In addition to the early response incentive, an incentive was used during the B&B:93/03 field test to reduce nonresponse among two groups: those who initially refused to participate in the study and those who could not be located but for whom a contact person could be reached. Sample members in the two 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.

A total of 85 sample members who initially refused to participate in the interview and another 20 who could not be located were eligible for the nonresponse incentive. Table 9 provides an overview of the contact and interview rates for these two nonrespondent groups. Overall, 61 percent of cases eligible for a nonresponse incentive were ultimately contacted, and 59 percent of those contacted were interviewed. Of the 85 initial refusals, 72 percent were contacted, and 57 percent of those contacted completed the interview. Of the 20 sample members initially coded as unable to locate, 3 (15 percent) were contacted for an interview, and all of those contacted completed the interview.

Table 9. B&B:93/03 contact and interview rates, by nonresponse incentive group

Incentive status	Total	Contacted		Interviewed	
		Number	Percent[1]	Number	Percent[1]
Total	925	748	80.9	673	90.0
Response incentive	105	64	61.0	38	59.4
Refusal incentive	85	61	71.8	35	57.4
Unable to locate incentive	20	3	15.0	3	100.0
No nonresponse incentive	820	684	83.4	635	92.8

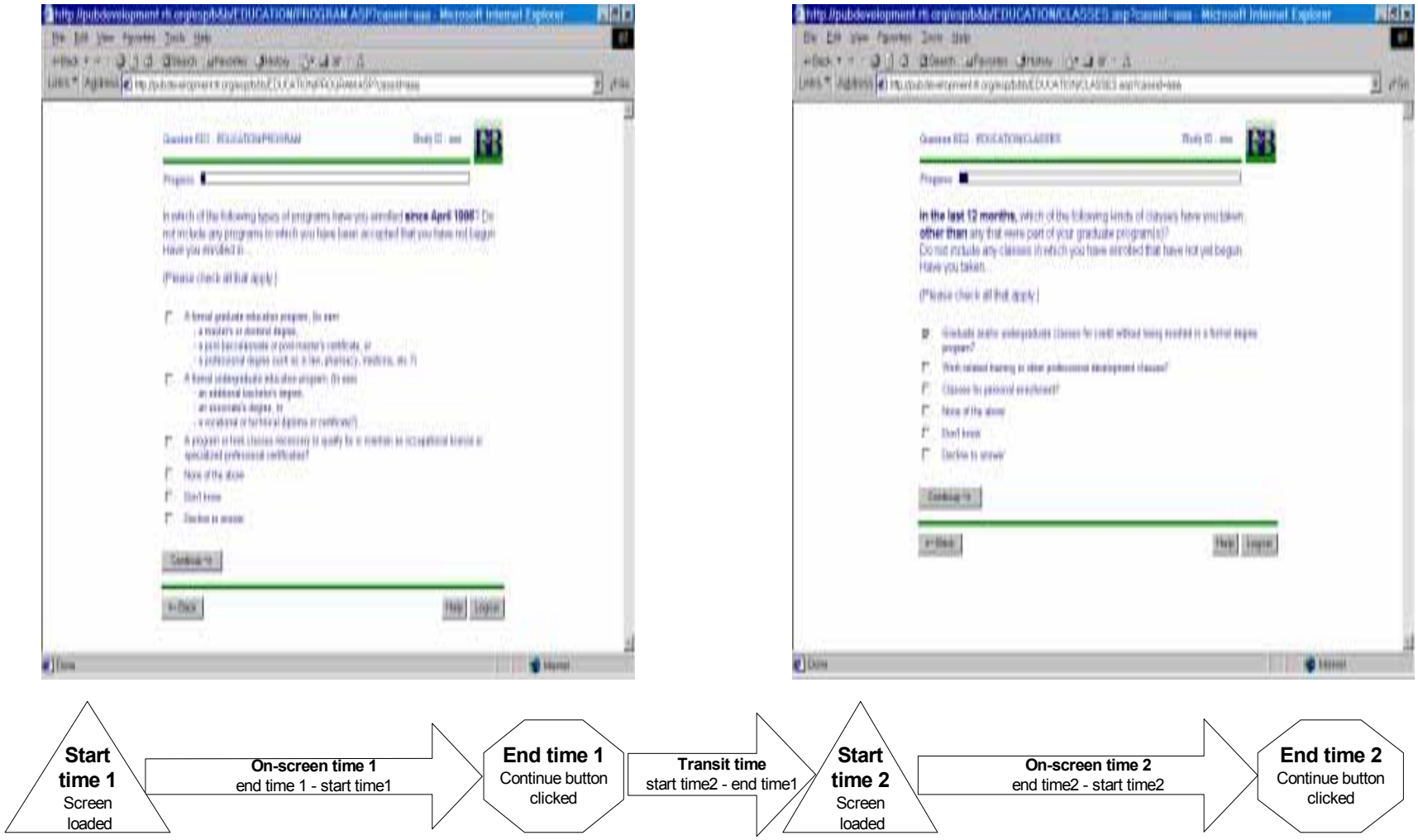
[1]Percent based on total within row under consideration.
NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

F. Interview Burden

The time burden associated with completion of the B&B:93/03 field test 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 3 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 that the respondent or interviewer clicked the "Continue" button to submit the answers from that page. Responses were transmitted to the server and new items were transmitted to the respondent's computer between interview screens.

Figure 3. 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).

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 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 10 presents the timing results (in minutes; combining on-screen and transition times) for the entire interview and by interview section, for all respondents. 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 almost 37 minutes.[6]

The employment section, which took an average of 11.4 minutes to complete, was among the longest sections in the interview. Questions in the employment section 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 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 and multiple bouts of unemployment.

The education section averaged over 10 minutes to complete. Since the education section collected all education experiences, many respondents were required to answer sets of questions within each subsection and for each school attended. Although few respondents enrolled in multiple undergraduate or graduate programs, the number of respondents enrolled for licensure or certification and for job training were much higher, increasing overall average time in the section.

Like the education section, the finance section averaged over 10 minutes, primarily because the finance questions required respondents to recall financial information such as salaries, total income, and loan amounts. They also had to estimate assets and debts by category. Anecdotally, respondents reported to telephone interviewers and through web comments that finance contained the most difficult questions to answer.

[6] Overall average time was calculated for all respondents. Overall time in the interview was also calculated separately for teachers, who answered all questions in Section C, and non-teachers who answered only three questions in Section C to determine their interest in teaching.

An entire section of the field test interview focused on questions for those who entered the teacher pipeline upon completion of the bachelor's degree in the 1991–92 school year, and any new entrants to the pipeline since 1991. Those respondents who have not taught and who have no interest in or plans for teaching were skipped around the teaching section after they answered the initial gate questions.

Table 10 also compares average times to complete the interview and to complete interview sections when the interview is self-administered and when the interview is administered by an interviewer, either by telephone or in person. Self-administered interviews averaged 40.5 minutes, longer than the 35.7 minutes to complete the interview with an interviewer (t=3.51, p<0.001). One reason for this difference, the large transit time experienced when the web interview was conducted by sample members using dial-up modems, is discussed further below.

Table 10. Average minutes to complete B&B:93/03 student interview, by interview section and mode of administration

	All respondents		Self-administered interviews		Interviewer-administered interviews	
Instrument section	Number of cases	Average time	Number of cases	Average time	Number of cases	Average time
Total interview[1]	652	36.7	225	40.5	427	35.7
Section A – Education	672	10.1	239	10.0	433	10.1
Section B – Employment	663	11.4	234	12.7	429	10.7
Section C – Teacher (teachers)	221	10.8	70	12.7	151	9.9
Section C – Teacher (non-teachers)	433	0.9	156	1.1	277	0.8
Section D – Background	654	4.6	226	5.0	428	4.5
Section E – Finance	652	10.3	225	8.7	427	11.1

[1] Total interview time combines on-screen and transit times across all sections and respondents, including teachers.

NOTE: A section was considered complete if the amount of time to complete the section was greater than zero and the section completion flag was set. One case was excluded from the analysis because of invalid timing data. Twenty respondents broke off the interview before completing all of the sections, so each section has a different number of cases. 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).

Self-administered interviews took longer than interviewer-administered interviews for the employment (t=4.80, p<0.0001), background (t=3.23, p<0.01), and teaching sections (t=4.28, p<0.0001).[7] It is not immediately obvious why these particular sections would take significantly longer to complete than the comparable CATI/CAPI sections when other sections of the survey did not. Possibly, respondents conducting the interview themselves spent additional time backing up to review prior responses and response options. CATI/CAPI interviews took longer in the finance section (t= –7.76, p<0.0001) when compared to self-administered interviews, most likely because of the time required to recall the amounts requested. Respondents administering the interview themselves could log out of it, collect the information required, then resume the interview

[7] For the teacher section, only the timing of those eligible to answer questions, that is, in the teacher pipeline, were compared.

to complete the section.[8] In addition, as discussed in chapter 4, self-administered interviews had higher rates of "don't know" and refusal responses in the finance section, and, therefore, their time in that section would have been less (see section B.1 of chapter 4).

Table 11 shows time in the interview divided into two values—on-screen time, which was the actual time spent reading and answering questions, and transit time, the idle time spent while pages were transmitted from the server to the local computer and loaded. CAPI respondents have been excluded from this analysis since the CAPI interview was housed locally on each interviewer's laptop, which made transit times virtually instantaneous. Overall, web respondents had a greater average total transit time (12 minutes) than did CATI respondents (4.4 minutes; t=13.15, p<0.0001). Web respondents also had longer section transit times for all five sections in the interview (education [t=8.20, p<0.0001]; employment [t=12.30, p<0.0001]; teaching [t=8.98, p<0.0001]; background [t=12.31, p<0.0001]; and finance [t=13.17, p<0.0001]). However, when transit time is removed from the total interview time, average interview time for web respondents (28 minutes) is statistically significant and *less* than for CATI respondents (31 minutes; t = –3.28, p<0.001).

Respondents who participated in the web debriefing (n=215) were asked which type of Internet connection they used to access the survey. Table 12 presents the average on-screen and transit times by Internet connection. Although dial-up via modem was by far the most common connection method, dial-up modem users took longer to complete the interview (F=20.69, p<0.0001) and had longer transit times (F=54.63, p<0.0001) than did any of the other connection methods.

G. Staff Burden and Effort

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 identification number, the type of problem, a description of the problem and resolution, its status (pending or resolved), and the approximate time it took to assist the caller.

[8]When web respondents broke off and then resumed an interview at a later time, the timer data showed a very large transit time between the last completed page and the first page of the later session. In most cases these were adjusted by flagging the large transit time and imputing a median transit time for the break-off event. Time that the respondent spent logged off of the interview was not included in transit time calculations.

Table 11. Average time (in minutes) on-screen and for data transit, by interview section, web respondents, and CATI respondents: 2003

Instrument section	All web and CATI respondents[1]			Web respondents			CATI[2] respondents		
	Number of cases	Average on-screen time	Transit time	Number of cases	Average on-screen time	Transit time	Number of cases	Average on-screen time	Transit time
Total interview	573	29.7	7.4	225	28.4	12.1	348	31.0	4.4
Section A – Education	592	8.5	1.6	239	7.4	2.6	353	9.2	0.9
Section B – Employment	583	8.8	2.5	234	8.8	3.9	349	8.9	1.5
Section C – Teacher (teachers)	206	8.8	2.0	70	8.9	3.8	136	8.8	1.1
Section C – Teacher (non-teachers)	369	0.5	0.4	156	0.6	0.5	213	0.6	0.2
Section D – Background	575	3.5	1.1	226	3.1	1.9	349	3.8	0.6
Section E – Finance	573	8.9	1.3	225	6.5	2.2	348	10.1	0.7

[1]Because field interviewers used stand-alone laptops for interviewing, transit times were negligible. Therefore, computer-assisted personal interviewing (CAPI) respondents have been excluded from analysis.

[2]CATI=Computer-assisted telephone interview.

NOTE: A section was considered complete if the amount of time to complete the section was greater than zero and the section completion flag was set. One case was excluded from the analysis because of invalid timing data. Twenty respondents broke off the interview before completing all of the sections, so each section has a different number of cases. 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 12. Average time (in minutes) on-screen and for data transit, by Internet connection speed: 2003

Internet connection speed	Number of cases[1]	Average on-screen time	Transit time
Dial-up modem	94	28.4	18.0
Fast connection	83	25.2	5.5
Cable modem	27	25.4	4.5
Digital subscriber line (DSL)	24	25.0	5.7
Integrated services digital network (ISDN)	3	25.8	3.8
Corporate local area network (LAN) (T1, T3, etc.)	29	26.1	6.4
Don't know	36	30.1	12.5
Decline to answer	2	15.3	5.6

[1] Of the 240 web respondents, 215 answered the debriefing questions at the end of the interview.

NOTE: Fast connection is the average interview time of respondents with a cable modem, digital subscriber line, ISDN, or corporate LAN. 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).

Help Desk staff were trained both to work the Help Desk hotline and to conduct telephone interviews when needed. Help Desk time was spent assisting sample members with questions about the web instrument and providing technical assistance to sample members who experienced problems while completing the self-administered web interview. 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. Only 102 calls were handled by the Help Desk during data collection.

Table 13 provides detail on the types of incidents encountered for cases that required Help Desk assistance. The majority of incidents (65 percent) recorded by the Help Desk were from sample members requesting their Study ID and/or password, with 12 percent of the calls asking about browser settings and computer problems. Program errors, reports of perceived logic problems, and reports of web site unavailability each accounted for 4 percent of Help Desk calls. Ten percent of calls were for other, miscellaneous issues.

Table 13. B&B:93/03 response pattern, by Help Desk incident type

Type of incident	Total incidents recorded	Percent of total incidents
Total	102	100.0
Study ID/password	66	64.7
Browser settings/computer problems	12	11.8
Program error	4	3.9
Routing/skip problems	4	3.9
Web site unavailable	4	3.9
Question about study/instrument	2	2.0
Other problems, not classifiable	10	9.8

NOTE: 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).

2. Interviewer Hours

Telephone interviewing for the field test required 1,097 telephone interviewer hours, exclusive of training, supervision, monitoring, administration, and quality circle meetings. The average time spent per completed interview was 3.11 hours. Since the average time to administer the interview was 35.7 minutes for CATI and CAPI cases, the large majority of interviewer time was spent in other activities. While a small percentage of non-interview 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.

3. Number of Calls

As indicated above, a significant amount of interviewer time was spent attempting to locate and contact sample members. Table 14 shows the number of telephone calls made to sample members overall, by mode of completion, and by prior response status.

Table 14. B&B:93/03 call counts, by interview status, mode of completion, and prior response status

Interview status	Number of cases	Number of calls	Average calls per case
Total	917	15,068	16.4
Interviewed	673	8,398	12.5
Web complete	240	1,837	7.7
CATI complete	353	5,094	14.4
CAPI complete	80	1,467	18.3
Not interviewed	244	6,670	27.3
Respondent status			
1997 Respondent	842	13,596	16.1
1997 Nonrespondent	75	1,472	19.6

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 total of 15,068 telephone calls were made during field test data collection, with an average of 16 calls made per sample member and a range of 0 to 96 calls, depending on response status and mode of completion. Those interviewed were called 13 times, on average, compared to those not interviewed, who were called an average of 27 times (t=11.26, p<0.0001). Interviews completed over the web required significantly fewer calls (8 calls) compared to CATI (14 calls, t=−6.94, p<0.0001) and to CAPI (18 calls, t=−7.02, p<0.0001). Sample members who were nonrespondents in 1997 were called an average of 20 times, compared to 16 times for those who were respondents in 1997, although this difference was not statistically significant (t=1.83, p=0.0682).

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 917 field test cases, 663 (72 percent) had at least one answering machine event. An average of 5 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 21 calls in cases in which an answering machine was reached at least once. Cases not reaching an answering machine (28 percent) required significantly fewer calls than those reaching an answering machine at least once (t=−18.57, p<=0.0001).

Among cases in which an answering machine was reached at least half of the time, it took an average of 22 call attempts to complete an interview, compared with 13 call attempts to complete interviews among cases in which an answering machine was reached less than half of the time. Cases with no answering machine events had a much lower rate of ever refusing (6 percent) than did cases with one or more answering machine events (17 percent, χ^2=19.0, p<0.0001). Similarly, cases with no answering

machine events had lower rates of a final refusal (1 percent) when compared to final refusals of cases with one or more answering machine events (4 percent, χ^2=3.8, p=0.05).

H. Conclusion

The purpose of the B&B:93/03 field test was to fully test all data collection procedures. The tracing, locating, and interviewing methods were successful for the field test and will be implemented again for the full-scale study. The results from the early response incentive experiment suggest that payment of an early response incentive does increase the likelihood of an early response. The full-scale study will implement a similar type of early response incentive to encourage web completion and reduce costs associated with telephone interviewing. To reduce respondent burden, several items will be eliminated or modified to decrease the overall time in the interview and to improve usability of the web instrument.

Chapter 4
Evaluation of Data Quality

The B&B:93/03 field test used a web-based instrument that could be either self-administered by sample members or administered to sample members by an interviewer. As a necessary step in preparation for the full-scale study, the B&B:93/03 field test included a number of evaluations to assess the quality of the data collected by the B&B instrument. 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. Implications of these results for the full-scale study are presented in chapter 5.

A. Usability of the Instrument

Developing a functional web survey for the B&B:93/03 field test 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. Several steps, therefore, 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.

1. Usability Testing

In designing the B&B instrument, commonly accepted standards were incorporated at the earliest stages of development (Dillman 2000). Once a fully integrated web instrument was available, small-scale testing was conducted to assess its usability. The primary focus of the usability testing was on the overall usability of the instrument. Prior to testing, several items were identified as being potentially difficult for sample members to complete. These included industry and occupation coding, as well as code assignment for the major field of study. In addition, the usability testing was designed to evaluate different screen layouts, item structures, and on-screen tools (e.g., help screens, progress indicators, and error messages).

Small-scale usability testing of the web-based instrument was conducted with eight participants recruited from the local area. Participants were recruited to reflect the B&B population, with each participant having completed college at some time during the 1991–92 school year. Four men and four women were interviewed. Seven of the

participants had completed college immediately after graduating from high school and were between the ages of 30 and 32 at the time that they took part in the testing. The remaining participant was about 45 years old. The participants were a racially diverse group.

All participants were interviewed at the same location, using the same computer, operating system, Internet provider, and connection speed. This consistency provided a controlled environment for completion of the interview, which eliminated the potential for differing participant reactions due to differences in the computing environment.[9]

Each participant completed the survey in the presence of a project staff member. Participants were instructed to "think aloud" as they entered their answers into the computer. As necessary, participants were prompted to articulate what they were doing as they entered responses, moved from question to question, accessed additional information, returned to an earlier question, and changed responses.

Overall, the application was determined to be fairly easy to work with and participants expressed very few concerns. The comments received from participants covered four main areas:

- Screen formatting/layout—On several screens it was not clear to the participant how to move forward to the next screen. In particular, this happened when participants were asked to answer two questions on one screen. These questions appeared in columns, with one question that asked the participant to pick one answer from the list and another that asked them to check all applicable responses. Most participants found screens of this type confusing.
- Font—Several participants reported that the font used throughout the survey was difficult to read, and others pointed out the need to use a bold typeface to emphasize certain words.
- Message boxes—Several participants reported that the message boxes did not clearly identify the mistake that they had made or how to fix it. Participants also noted that the message text was written using overly technical terminology.
- Text boxes—Participants pointed out that, in some situations, text boxes were not long enough to accommodate their entire answer. In other cases, the boxes were larger than the allowable typing space, which created additional confusion.

Based on the results of the small-scale testing, a number of revisions were made to the instrument to improve its usability by reducing the complexity of the response tasks and clarifying the nature of the data requested.

[9]Testing to ensure that the web-based instrument operated equally well under different operating systems, including Windows, Mac, and Linux, was conducted early in the development process.

2. Interview Progress

An experiment was embedded within the web self-administered interview[10] to see which of two types of respondent *motivators* would be the most successful in encouraging sample members to complete a full interview. All 925 field test sample members were randomly assigned to receive either a text message or a progress bar indicator. The two types of motivators were compared to see which would result in fewer missing data and shorter average interview times.

The primary intent of the text message was to thank the respondent for the information that he/she had already provided and encourage him/her to continue with the questionnaire until all questions had been answered. The idea for using text messages to motivate respondents was adapted from research conducted by Charles Cannell in the 1970s on methods for encouraging survey respondents to provide thorough and accurate responses (Cannell, Miller, and Oksenberg 1981; Cannell, Oksenberg, and Converse 1979). A similar model might be effective for ensuring that web survey respondents carefully attend to the response task and not exit the interview before they have completed all the questions.

The progress bar was also designed to motivate respondents to complete the interview. The intent was to visually indicate to respondents how much of the interview had been done, and how much remained to be completed. The progress bar was rectangular in shape and located at the top of each interview screen. The bar was displayed for the duration of the interview. As the respondent progressed through the interview, the bar lengthened so that the respondent could determine the proportion of the section that had been completed.

To calculate indeterminate rates, a composite score of "don't know" and refusal responses was created for selected variables in the survey. Selected variables included key content items (such as pursuit of additional education) and opinion and fact questions (such as evaluating importance of education versus martial status). A total of 62 variables were evaluated. There was no difference by type of motivator in the instance of missing data. There were too few cases in each treatment group to evaluate the impact of break-off rates on data quality with respect to the presence of a text message versus a progress bar.

There was also no statistically significant difference (χ^2=1.6, p=0.1) in the average interview length for respondents receiving the text message, who averaged 37.9 minutes to complete the interview, and respondents receiving the progress bar, who averaged 36.2 minutes to complete the interview. Given the measures evaluated, therefore, it does not appear that the outcomes of missing data and interview time are differentially affected by the type of respondent motivator used.

[10]CATI/CAPI cases were excluded from this analysis.

3. 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). Also included on every help text screen was a toll free number to the B&B:93/03 Help Desk. Counters were used to determine the number of times that each help screen was accessed. Items with high rates of help text access indicated that web respondents or CATI/CAPI interviewers needed additional information about the question before giving a valid response.

Overall, less than 1 percent of respondents used help text at any point in the interview. CATI/CAPI interviewers called help text for 287 (1.1 percent) of the 26,135 screens accessed during the field test data collection period, compared to 86 (0.6 percent) of 14,590 screens accessed by self-administered respondents. Although help text usage was very low, the observed difference in rates is statistically significant. Interviewers accounted for the majority of accesses to help text, most likely due to the fact that they are both trained and encouraged to use the help text whenever needed.

The question "In your February job as a [job title], did your employer provide you with any of the following benefits?" had the highest number of help text accesses, with a total of 16 out of the 536 times it was administered. The reason for the relatively high number of accesses may be attributed to the inclusion of benefits, such as a transit subsidy or flexible spending account, which may have been unfamiliar to some respondents. The help text for this item contained a thorough description of each benefit.

Other confusing questions that were identified by high counts of help text accesses and rates of indeterminate responses included a set of finance questions that requested a dollar value of assets ("Please give a dollar amount for each of the following. . ."), an item that asked about telecommuting ("In your February job, did your employer allow you to telecommute?"), and an item that asked about continuing education units ("Did you earn any Continuing Education Units [CEUs] for any of the courses you've taken in the last 12 months?"). Respondents may not have been clear on how to calculate the dollar value of their assets, or may not have been familiar with the terms "telecommute" and "Continuing Education Units."

A comparison of help text accesses by mode revealed that web respondents and CATI/CAPI interviewers accessed help text at essentially the same rate, nearly 13 percent for web respondents and approximately 12 percent for CATI/CAPI interviewers.

4. Coding Systems

The B&B field test instrument included tools that allowed online coding of literal responses for major/field of study, area of licensure or certification, occupation, industry, postsecondary institutions attended, and, for teachers, the elementary and/or secondary schools at which the respondent has been employed. When administered by interviewers,

these online coding systems improve data quality by capitalizing on the availability of the respondent to clarify coding choices at the time the coding was performed; interviewers were trained to use probing techniques in assisting the online coding process. When self-administered by B&B sample members, help text and limited supporting text on screen were available to assist online coding. However, the extent to which the quality of self-administered online coding would be improved or compromised relative to interviewer-administered results was unknown.

The final field test data file contained both the literal strings and selected codes for field of study, licensing/certification, occupation, and industry responses, allowing recoding by a coding expert to be easily included in field test data evaluation activities and comparisons made across interview modes. The first step in the analysis was to examine a 10 percent sample of each set of coding results. Expert coders evaluated the verbatim strings for completeness and for the appropriateness of the assigned codes, determining whether a different code should be assigned or if a string was too vague to code.

Table 15 shows the results of the 10 percent recode analysis for each coder: major, licensing/certification, occupation, and industry. Overall, no statistically significant mode differences were detected in the coding results—expert coders agreed with web respondent coding (92.2 percent) at about the same rate as they agreed with CATI/CAPI interviewer coding (88.8 percent; z=0.62, p>0.10). Table 15 also provides agreement rates by mode for each type of coding system used. Across both modes of interview, there were no statistically significant differences in the rate of agreement between respondents or interviewers and the expert coders suggesting that the coding systems were used with comparable facility by both self-administered and interviewer-administered respondents.

Table 15. Summary of 10 percent recode analysis for online coding systems: 2003

	Self-administered		CATI[1]/CAPI[2]	
Type of coding	Coding attempts sampled	Percent original code correct	Coding attempts sampled	Percent original code correct
Total	51	92.2	89	88.8
Major	‡	‡	‡	‡
License/certificate	6	100.0	16	81.3
Occupation	19	94.7	31	80.7
Industry	22	86.4	35	97.1

‡ Too few cases to report.
[1]CATI=Computer-assisted telephone interview.
[2]CAPI=Computer-assisted personal interview.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

Use of the "other" response in the interview coding systems was also examined following data collection. No statistically significant difference was detected in the rate at which web respondents used the "other" response for the major, licensing/certification, occupation, and industry coding systems (3.1 percent) compared to CATI/CAPI

respondents (4.6 percent; see table 16). The associated strings were recoded at about the same rates. A similar analysis evaluated use of the "other" response in the postsecondary and elementary/secondary school coding systems described above. Again, web and CATI/CAPI rates for use of the "other" option in the school coders were similar and could be recoded as schools already contained in the database.

Throughout the field test interview, "other, specify" with a text string was available as a response when the options provided in a list were not appropriate or applicable. Following data collection, expert coders examined each of the occurrences of the "other, specify" responses to determine if the associated text strings could be recoded into existing response categories. Table 16 presents the recoding results by mode of interview. In comparing web and CATI/CAPI responses, use of the "other, specify" actually differed by mode, with web respondents selecting "other, specify" more often (18.6 percent of all responses) than interviewers (16.4 percent; $p<0.05$), but the rate at which the text strings could be recoded did not differ by mode.

Table 16. Number and percent of "other, specify" responses provided during interviewing, and rate of recoding to existing categories: 2003

	Web			CATI[1]/CAPI[2]		
Type of coding	Number of responses	Percent responding "other"	Percent of "other" responses recoded	Number of responses	Percent responding "other"	Percent of "other" responses recoded
Coding systems[3]	517	3.1	31.3	743	4.6	32.4
School coding	183	13.1	25.0	365	16.1	28.8
Questions with "other, specify"	2,066	18.6	26.3	4004	16.4	26.7

[1]CATI=Computer-assisted telephone interview.
[2]CAPI=Computer-assisted personal interview.
[3]Coding systems include the major, licensing/certification, occupation, and industry coders.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

B. Effectiveness of the Instrument

For the B&B field test, 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. In addition, a reinterview was conducted with a subsample of respondents to assess the temporal stability of the data. Finally, a comparison of responses for selected items evaluated the extent to which responses varied across modes. The results of each of these evaluations are presented below.

1. Indeterminate Responses

Every item in the interview accommodated two different types of nonresponse, "don't know" and refusal responses. Refusal responses to interview questions were most common for items considered sensitive by respondents. "Don't know" responses may

result from a number of potential circumstances, the most obvious reason being that the answer is truly unknown to the respondent. "Don't know" responses may also be evoked when (1) question wording is not understood by the respondent, (2) the respondent is hesitant to provide a "best guess" response, or (3) the respondent implicitly refuses to answer a question. Refusal and "don't know" responses introduce indeterminacies in the data set. Indeterminate rates were calculated overall and by mode. Web respondents provided an indeterminate response 5 percent of the time; CATI/CAPI respondents provided an indeterminate response 2.5 percent of the time.

Table 17 lists items with over 10 percent indeterminate responses. Item nonresponse rates were calculated based on the number of sample members for whom the item was applicable and asked. A discussion of item nonresponse by section and by mode of data collection is presented below.

a. Education and employment

Rates of indeterminacy were relatively low in both the education and employment sections. In the education section, the month and year that respondents expect to earn a certificate resulted in a high number (32 percent) of "don't know" responses. Of the items in the employment section, questions regarding income resulted in a high number of indeterminate responses, mostly from refusal responses. Respondents provided an indeterminate response about 17 percent of the time for their current salary and nearly 11 percent of the time for their February salary. No difference was found in the overall rate of indeterminacy between web and CATI/CAPI respondents for the education or employment section.

b. Teaching

In the teaching section, teachers were asked about their income from the district as well as other sources, and all of these items had high rates of indeterminacy. Nearly 20 percent of teachers refused or did not know their current base-year salary. Teachers were reluctant to provide their nonschool income (12 percent) and the other income that they earn from the district in which they teach (11 percent). The set of questions pertaining to future career plans in the teacher section also resulted in a high percentage of "don't know" responses. Nearly 12 percent of teachers did not know whether they planned to move to a nonteaching education job, and 10 percent were unsure if they would continue teaching in the future. For these items, of course, "don't know" was a reasonable response.

Table 17. B&B:93/03 interview item nonresponse for items with more than 10 percent "don't know" or "refused"

Section and variable name	Label	Number asked	Percent don't know	Percent refused	Combined percent
Education					
B3CTXM1	Month expect to earn certificate	47	27.7	4.3	31.9
B3CTXY1	Year expect to earn certificate	47	27.7	4.3	31.9
Employment					
B3CURSAL	Salary for current job	52	1.9	15.4	17.3
B3CURUNT	Unit of time for salary in current job	52	1.9	15.4	17.3
B3FEBSAL	Salary for February job	579	1.6	9.2	10.7
B3FEBUNT	Unit of time for salary in February	579	1.6	9.2	10.7
Teachers					
B3SALR1	Current position base annual teaching salary	221	14.5	5.0	19.5
B3TSY1	Teaching school county	45	15.6	2.2	17.8
B3OOIN1	Nonschool income	221	6.8	5.4	12.2
B3OINC1	Other income from district	221	6.8	4.5	11.3
B3JBPLN	Plan to move to nonteaching education job	113	11.5	#	11.5
B3TCHFTR	Plan to continue teaching	221	10.0	0.5	10.4
Finances					
B3ALINS	Value of life insurance	653	23.3	21.4	44.7
B3ARETIR	Value of retirement funds	653	16.1	24.8	40.9
B3AINVST	Value of investments	653	12.3	25.6	37.8
B3ACOLL	Value of collectibles	653	11.3	22.5	33.8
B3ACASH	Amount of cash on hand	653	4.9	26.3	31.2
B3AAUTO	Value of automobiles	653	7.0	19.9	27.0
B3DHOME	Amount of home mortgage	653	5.8	18.5	24.3
B3AHOME	Value of home	653	4.7	19.1	23.9
B3DAUTO	Amount owed on automobiles	653	4.9	15.3	20.2
B3OINC02	Total household income from all sources	653	4.9	14.5	19.4
B3DOTLOA	Amount owed on other types of loans	653	3.8	15.2	19.0
B3CREDT	Amount of credit card balance	653	3.1	15.5	18.5
B3INS02	Spouse/partner income from work in 2001	478	3.1	13.0	16.1
B3INC02	Income from work in 2001	653	2.6	10.9	13.5
B3MTGAMT	Monthly rent or mortgage payment	653	0.8	10.7	11.5
B3SRPOM	Month that spouse/partner repaid loans	55	60.0	1.8	61.8
B3SRPOY	Year that spouse/partner repaid loans	55	49.1	3.6	52.7
B3SRPBM	Month that spouse/partner began repaying loans	76	32.9	1.3	34.2
B3RPYOM	Month that respondent repaid loans	156	25.0	3.2	28.2
B3SRPBY	Year that spouse/partner began repaying loans	76	25.0	1.3	26.3
B3SRPAMT	Spouse's/partner's monthly student loan payment	76	18.4	5.3	23.7
B3SEDLN	Amount spouse/partner borrowed in loans for education	478	16.7	3.8	20.5
B3SRPST	Spouse/partner currently repaying student loans	166	6.0	7.2	13.3
B3RPYOY	Year that respondent repaid loans	156	9.6	3.2	12.8
B3UGLN	Total amount undergraduate loan	555	7.4	4.0	11.4
B3SRPTP	Spouse's/partner's type of repayment plan	76	10.5	#	10.5

#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).

One explanation for the high rate of "don't know" responses is related to the mode of data collection. Table 18 presents indeterminate responses for the teacher section by mode of data collection. Web respondents had a higher refusal rate (10 percent) when asked about their nonteaching income than their CATI/CAPI counterparts (3 percent, χ^2=4.17, p<0.05). Web respondents also had a higher indeterminate rate (21 percent) for their current other income from the district in which they teach than did CATI/CAPI respondents (7 percent, χ^2=10.5, p<0.005). Web respondents were more likely to express uncertainty about whether they plan to move to a nonteaching education job (χ^2=21.8, p<0.0001) and whether they plan to continue teaching (χ^2=8.5, p<0.005) than CATI/CAPI respondents.

c. Finance

Close to 25 percent of the finance questions resulted in 10 percent or more indeterminate responses. Some of the information required for this section could not be readily recalled. For example, over 23 percent of respondents did not know the value of their life insurance, retirement funds, investments, and collectibles. Based on comments made in the debriefing section, respondents considered many of the questions in the finance section to be very sensitive and intrusive. As a result, many respondents refused to answer questions that pertained to their assets and debts. The value of life insurance, retirement funds, investments, collectibles, and cash on hand were refused by over 20 percent of respondents.

Like in the teaching section, the high refusal rates for the finance questions may be due in part to the mode of data collection (see table 19). Web respondents were more likely to respond with an indeterminate response than CATI/CAPI respondents. All of the questions regarding assets and debts resulted in web respondents having a significantly higher percentage of refusals as well as a higher overall rate of indeterminacy. For example, web respondents were more likely to refuse to provide the value of their life insurance (36 percent, χ^2=19.9, p<0.0001), and they were more likely to provide an indeterminate response (57 percent, χ^2=42.6, p<0.0001) than were CATI/CAPI respondents (14 percent and 38 percent, respectively). With one exception, there was no statistical difference in the use of the "don't know" option for the asset/debt items when comparing web and CATI/CAPI respondents.

The set of items asking the respondent about their spouse's/partner's student loans evoked a high number of "don't know" responses. Many respondents did not know the month (60 percent) and year (49 percent) that their spouse/partner repaid their student loans. Web respondents were also more likely to provide an indeterminate response to their spouse's/partner's income (22 percent, χ^2=5.8, p<0.05) and their income from work (18 percent, χ^2=5.3, p<0.05) when compared with CATI/CAPI respondents (13 and 16 percent, respectively). Furthermore, web respondents provided a higher percentage of indeterminate responses to questions pertaining to their spouse's/partner's education loans (χ^2=21.7, p<0.0001) and their own undergraduate student loans (χ^2=16.0, p<0.0001).

Table 18. B&B:93/03 interview item nonresponse for items in the teacher section with more than 10 percent "don't know" or "refused," by mode

Variable name	Label	Number asked		Percent don't know		Percent refused		Total percent indeterminate	
		Web	CATI[1]/ CAPI[2]	Web	CATI/ CAPI	Web	CATI/ CAPI	Web	CATI/ CAPI
B3OINC1	Current: other income from district	70	151	10.0	5.3	11.4	1.3	21.4	6.6
B3OOIN1	Current: nonteaching income	70	151	8.6	6.0	10.0	3.3	18.6	9.3
B3JBPLN	Plan to move to nonteaching education job	33	80	33.3	2.5	#	#	33.3	2.5
B3TCHFTR	Plan to continue teaching	70	151	18.6	6.0	1.4	#	20.0	6.0

#Rounds to zero.
[1]CATI=Computer-assisted telephone interview.
[2]CAPI=Computer-assisted personal interview.
NOTE: A chi square statistic was not calculated for variables in which there were small (n<6) cell sizes.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

Table 19. B&B:93/03 interview item nonresponse for items in the finance section with more than 10 percent "don't know" or "refused," by mode

Variable name	Label	Number asked		Percent don't know		Percent refused		Total percent indeterminate	
		Web	CATI[1]/ CAPI[2]	Web	CATI/ CAPI	Web	CATI/ CAPI	Web	CATI/ CAPI
Assets									
B3ALINS	Value of life insurance	226	427	20.8	24.6	35.8	13.8	56.6	38.4
B3ARETIR	Value of retirement funds	226	427	10.6	19.0	35.8	19.0	46.4	38.0
B3AINVST	Value of investments	226	427	11.5	12.6	36.3	19.9	47.8	32.5
B3ACOLL	Value of collectibles	226	427	14.6	9.6	35.8	15.5	50.4	25.1
B3AAUTO	Value of automobiles	226	427	7.5	6.8	32.7	13.1	40.2	19.9
B3ACASH	Amount of cash on hand	226	427	3.1	5.9	37.2	20.6	40.3	26.5
B3AHOME	Value of home	226	427	4.0	5.2	33.6	11.5	37.6	16.7
Debts									
B3DHOME	Amount of home mortgage	226	427	5.8	5.9	28.3	13.3	34.1	19.2
B3DAUTO	Amount owed on automobiles	226	427	5.3	4.7	25.2	10.1	30.5	14.8
B3DOTLOA	Amount owed on other types of loans	226	427	8.0	1.6	25.2	9.8	33.2	11.4
B3DCREDT	Amount of credit card balance	226	427	4.0	2.6	24.3	10.8	28.3	13.4
Other finance items									
B3OINC02	Total household income from all sources	226	427	7.1	3.7	19.5	11.9	26.6	15.6
B3INS02	Spouse/partner income from work in 2001	172	306	4.1	2.6	17.4	10.5	21.5	13.1
B3INC02	Income from work in 2001	226	427	3.1	3.7	14.6	11.9	17.7	15.6
B3SRPAMT	Spouse's/partner's monthly student loan payment	25	51	32.0	11.8	12.0	2.0	44.0	13.8
B3SEDLN	Amount spouse/partner borrowed in loans for education	172	306	26.7	11.1	5.2	2.9	31.9	14.0
B3SRPST	Spouse/partner currently repaying student loans	73	93	6.8	5.4	12.3	3.2	19.2	8.6
B3UGLN	Total amount undergraduate loan	192	363	12.5	4.7	6.3	2.8	18.8	7.5

[1]CATI=Computer-assisted telephone interview.
[2]CAPI=Computer-assisted personal interview.
NOTE: A chi square statistic was not calculated for variables in which there were small (N<6) cell sizes. 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)

d. Comparable rates of CATI/CAPI indeterminacy

Overall indeterminacies in the B&B field test interview were somewhat higher when compared to similar studies with similar populations. To understand this difference, web respondents were excluded from an item-level comparison with CATI/CAPI respondents to other studies (NPSAS:1996, B&B:93/97, and the 1996/01 Beginning Postsecondary Students Longitudinal Study [BPS:96/01]). With one exception, no statistically significant difference was detected in item-level rates of indeterminacy (see table 20). The fact that the interview was offered on the web for the first time may be the primary reason for higher than normal indeterminate rates in the B&B interview and specifically in the finance section.

Table 20. Comparison of indeterminate rates across studies: 2003

Variable label	B&B:93/03 percent indeterminate	Other study percent indeterminate
Amount of cash on hand	26.5	21.0[1]
Amount of home mortgage	19.2	17.0[1]
Value of home	16.6	20.9[1]
Spouse/partner income from work	13.1	16.7[2]
Income from work	11.2	11.3[2]
Amount spouse/partner borrowed in loans for education	14.1	24.8[2]
Spouse's/partner's monthly student loan payment	13.7	11.8[3]

[1] Source: U.S. Department of Education, National Center for Education Statistics, 1996 National Postsecondary Student Aid Study: (NPSAS:96).
[2] Source: U.S. Department of Education, National Center for Education Statistics, 1996/01 Beginning Postsecondary Students Longitudinal Study (BPS:96/01).
[3] Source: U.S. Department of Education, National Center for Education Statistics, 1993/97 Baccalaureate and Beyond Longitudinal Study (B&B:93/97).
NOTE: Web respondents have been excluded because prior studies did not offer a web, self-administered option.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

2. Break-offs

Of the 673 interviews conducted during the B&B field test, only 20 interviews were not completed through the end of the last section, that is, were partial interviews. To be considered a partial interview, the respondent had to answer questions at least through the end of the education section, the first section in the interview. Among the 20 partial interviews, most of the break-offs occurred either in the second section on employment (nine cases) or the third, teacher section (seven cases). Of the seven cases that ended in the teacher section, four broke off at the gate question that determines whether a respondent qualifies to answer questions in the teaching section. Otherwise, no specific pattern of interview break-offs occurred.

3. Reliability of Responses

During instrument development for the B&B:93/03 field test study, a short reinterview was developed to assess the short-term temporal stability of key interview items. The reinterview was designed to target items that were newly designed for the B&B:93/03 interview. It also provided an opportunity to test for differences across mode of administration—i.e., to determine if the temporal stability of responses was the same for those who completed the interview via

CATI and those who completed the interview themselves over the web. Respondents were reinterviewed in the same mode as the original interview was completed.

A randomly selected subsample of 215 interview respondents (111 CATI and 104 self-administered via the web) was asked to participate in the reinterview process. From this group, 175 reinterviews were completed, for an overall reinterview response rate of 81.4 percent. Reinterview response rates for CATI and web respondents were 82.0 and 81.0, respectively. The reliability statistics presented in this section are based on these 175 respondents. Sample member recontacting took place at least 3 weeks after the initial interview.

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. Responses in the initial interview and the reinterview were then compared using two measures of temporal stability for all paired responses. The first, *percent agreement*, determined the percentage of reinterview responses that exactly matched the original responses from the main interview. The second measure evaluated temporal stability using either Kendall's tau-*b* (τ_b) or Cramer's V. Kendall's tau-*b*, which takes into account tied rankings (Agresti 1984; Kendall 1945), was used for questions that were answered using ordered categories (e.g., very important, somewhat important, and not important). Cramer's V was used for items with discrete, unordered response categories (e.g., yes/no responses). Lack of agreement or a low relational statistic value (typically below 0.60) for responses collected at two time points reflects instability over time due to measurement error.

While analyses were based on the 175 respondents who completed reinterviews, effective sample sizes are presented for each item because analyses were further restricted to cases with determinate responses to the relevant items in both interviews. Because not all items were applicable to all respondents (e.g., only respondents who attended graduate school were asked the items about the importance of their graduate education), variation exists in the number of cases on which the reliability indices were based.

a. Education-related items

The first set of reinterview items was selected from the education section of the main interview. The first question asked respondents to identify the highest level of education that they expect to attain in their lifetime. The second involved a set of items asking how important—very, somewhat, or not important—each aspect of their undergraduate education is 10 years later. The results of the reinterview analysis for these two items are presented in table 21. The third and final set of education items asked respondents to rate the importance of an undergraduate education to specific aspects of their life now. The results of the reinterview analysis for this item are presented in table 22.

The first question about highest expected level of education had good response stability: percent agreement was almost 80 percent, and the relational statistic was high at 0.83. Further examination of main interview and reinterview responses to this question revealed that the response categories of postbaccalaureate certificate and post-master's certificate were primarily the source of the temporal instability at reinterview. Since these are relatively new degree offerings, respondents may be unclear of the differences. In the full-scale interview, the help text will be revised to further clarify all degree types.

Table 21. B&B:93/03 reliability indices for education items: Highest level of education expected and importance of undergraduate education 10 years later

Item series	Number of cases[1]	Percent agreement[2]	Relational statistic[3]
Highest level of education expected	152	79.6	0.83
How important is each of the following aspects of your undergraduate education to your life now, 10 years later?			
Undergraduate major(s)	172	77.3	0.67
Liberal arts courses taken	172	60.5	0.50
Professional courses taken	164	62.8	0.40
Quality of instruction	174	69.5	0.44
Relationship with faculty	171	56.7	0.52
Internship and other work	141	66.7	0.61
Support services	161	61.5	0.52
Social contacts	170	64.1	0.55
Extracurricular activities	163	69.3	0.60

[1]Analyses were conducted only for respondents with determinate responses on both the initial interview and the reinterview; not all questions were applicable to all respondents.

[2]Percent agreement is based on an exact match for nominal and ordinal measures. For continuous variables, values had to be within one standard deviation.

[3]Relational statistic used was Kendall's Tau, τ_b.

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

The series of items asking about the degree of importance of various undergraduate experiences is new to the B&B study, and it had generally low indicators of reliability. Percent agreement for these items ranged from about 57 to 77 percent, with relational statistics ranging from 0.40 to 0.67. Only the item asking the importance of the undergraduate major was reasonably stable over time (77 percent agreement and a relational statistic of 0.67).

A cross tabulation of the main and reinterview responses shows that most ratings concentrated among the "very" and "somewhat important" categories, while most of the temporal instability was among those who initially rated an item as "not important." This response instability among cases reporting "not important" decreases the relational statistic, because it results in an unbalanced distribution.

For a few of the items—specifically, relationship with faculty and social contacts—initial ratings were relatively evenly distributed among the categories of "very," "somewhat," and "not important." However, percent agreement for both items was low—56.7 and 64.1 percent, respectively.

Table 22. B&B:93/03 reliability indices for education item: Importance of undergraduate education as preparation for life 10 years later

Item series	Number of cases[1]	Percent agreement[2]	Relational statistic[3]
How important was your undergraduate education in preparing you for each of the following aspects of your life now, 10 years later?			
Work and career	174	70.7	0.47
Further education	168	69.6	0.51
Raising a family	148	54.7	0.43
Becoming established in a community	167	62.9	0.52
Financial security	171	61.4	0.42
Taking on new challenges	172	69.2	0.48
Making informed choices	173	67.1	0.48

[1]Analyses were conducted only for respondents with determinate responses on both the initial interview and the reinterview; not all questions were applicable to all respondents.

[2]Percentage agreement is based on an exact match for nominal and ordinal measures. For continuous variables, values had to be within one standard deviation.

[3]Relational statistic used was Kendall's Tau, τ_b.

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

The next question set asked respondents to rate the importance of their undergraduate education to various aspects of their life now. Results are presented in table 22. Overall, the two reliability indicators for this set of items suggest poor temporal stability in the responses. Percent agreement ranges from 55 to 71, with the relational statistic ranging from 0.42 to 0.52.

As with the first rating question set discussed above, the response pattern showed a concentration of "very important" and "somewhat important" ratings, with most of the disagreement between interview and reinterview occurring among the cases who initially gave ratings of "not important." Cases of nonagreement between the initial interview and reinterview tended to rate the importance of their undergraduate education higher in the reinterview—those who initially gave ratings of "somewhat" and "not important" tended to upgrade their response to "very" or "somewhat important" by the reinterview. For example, 62 percent of respondents initially rated their undergraduate education as "very important" in preparing them for work and career, 36 percent rated it as "somewhat important," and only 3 percent rated it as "not important." On reinterview, only 40 percent of those who initially gave a rating of "not important" gave the same response, while 81 percent of those who initially rated it as "very important" gave the same response.

In addition to comparing temporal stability within mode of interview completion, percent agreement and relational statistic values were compared across modes to determine if the same items yield equivalent levels of stability. Very few differences were identified among the education items. Importance of liberal arts courses taken showed percent agreement that was higher among the self-administered web interviews (73.5 percent) than among CATI interviews (48.3 percent; χ^2=11.39, p<0.001). The difference in the values of the web and CATI relational statistic (0.62 and 0.41) was also statistically significant (t= –2.02, p<0.05).

When compared across modes, the item rating the importance of the undergraduate education to taking on new challenges showed no statistically significant difference in the percent agreement for web and CATI respondents (75.6 and 63.3 percent, respectively; p=0.08).

However, the relational statistic comparing the two modes was statistically significant (0.38; $t = -2.02$, $p<0.05$). Consequently, while the overall rate of nonagreement between the two modes was the same, the distribution of responses from initial to reinterview was different. Among the respondents who originally rated their undergraduate education as "very important" in preparing them to take on new challenges, 83 percent of web respondents gave the same rating during the reinterview, while 58 percent of CATI respondents did so. No other statistically significant differences in percent agreement or the relational statistic were observed for this set of items.

b. Employment-related items

Table 23 presents the results of reliability analyses for employment questions about the job held in February. Overall reliability for these items was quite good, with percent agreement ranging from 79 to 89 percent and the relational statistics ranging from 0.68 to 0.83. Results for the item asking the number of times out of work for at least 3 months, while not the most reliable in the series, were quite strong (84 percent agreement and a relational statistic of 0.79), considering that respondents were asked to recount their employment history back to 1996.

The item with the lowest indicators of reliability—the level of flexibility in the job held in February—had different results when administered as a web and CATI interview. Agreement for web reinterviews (72 percent) was lower than for CATI interviews (86 percent; $\chi^2=3.98$, $p=0.46$), while the relational statistic was not statistically significantly different across modes (0.66 for web and 0.86 for CATI; $t=1.67$).

Table 23. B&B:93/03 reliability indices for employment items: Description of employer, flexibility of job, hours spent at work, and time out of work

Item series	Number of cases[1]	Percent agreement[2]	Relational statistic[3]
Employment			
Description of February employer	144	82.6	0.83[4]
Flexibility of February job	131	78.6	0.77
Greatest number of hours spent in the office	148	88.5	0.68
Times out of work 3 months or more	172	83.7	0.79

[1]Analyses were conducted only for respondents with determinate responses on both the initial interview and the reinterview; not all questions were applicable to all respondents.

[2]Percentage agreement is based on an exact match for nominal and ordinal measures. For continuous variables, values had to be within one standard deviation.

[3]Unless otherwise noted, the relational statistic used was Kendall's Tau, τ_b.

[4]Relational statistic used was Cramer's V.

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

Web responses to the item that asks where the respondent spends the greatest number of hours while working (in the office, at a job site or other location, telecommuting, or other) showed higher reliability in terms of the relational statistic (0.79) than did CATI responses (0.53; $t = -2.32$, $p<0.05$), but the rate of agreement within web and CATI did not differ statistically. For both modes of administration, the majority of cases responded that they spent the greatest number of hours in the office (98 percent for web and 95 percent for CATI). Among those who initially reported working at a job site or other location, 29 percent of web interviews gave the same response during the reinterview, compared with 71 percent among CATI interviews.

In another question series new to the B&B study, respondents were asked about the importance of various types of education and training to their current job. Measures of reliability are shown in table 24, and the results are mixed. For instance, experience on the job has the highest percent agreement of all items examined (91 percent), yet the relational statistic for this item is low (0.33) because so few cases reported that it was not important. Ninety-four percent of all cases said that experience on the job was very important to their current job, and 94 percent of those cases provided the same response during the reinterview. Only 6 percent of those who participated in the reinterview initially reported that experience on the job was somewhat important, and of those, 44 percent maintained the same response, while 56 percent changed their answer to very important during the reinterview.

In contrast, any other education (excluding undergraduate and graduate education) and experience from other jobs had relatively low indicators of temporal stability, perhaps because these categories are very broad and lack sufficient boundaries. For use in future studies, these items would be improved if they refer to a specific educational program and a specific job.

Table 24. B&B:93/03 reliability indices for employment items: Importance of prior education and experience to current job

Item series	Number of cases[1]	Percent agreement[2]	Relational statistic[3]
How important are each of the following to your current job?			
Undergraduate education	144	77.8	0.66
Any other education	123	62.6	0.43
Formal on-the-job training	133	69.2	0.41
Experience on the job	145	91.0	0.33[4]
Experience from other jobs	132	67.4	0.56

[1]Analyses were conducted only for respondents with determinate responses on both the initial interview and the reinterview; not all questions were applicable to all respondents.

[2]Percentage agreement is based on an exact match for nominal and ordinal measures. For continuous variables, values had to be within one standard deviation.

[3]Relational statistic used was Kendall's Tau, τ_b.

[4]The relational statistic is deceptively deflated due to insufficient variation across valid response categories. As a result, minor changes on the distribution of responses between the original and reinterview significantly lower the correlation coefficient.

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

c. Finance items

Results of reliability analyses concerning finance items are presented in table 25. Respondents were asked the number of dependents that they claimed when they filed their 2001 taxes. Overall percent agreement was fairly low at 69 percent, with a moderate relational statistic, 0.79. Examination of the data from the initial interview and reinterview revealed that agreement was highest among cases who initially reported claiming no dependents or only one dependent (80 and 81 percent, respectively). While claiming two dependents was the most common response during the initial interview, percent agreement was only 66 percent for this group. Help text should be written to clearly define the term dependent as it is used for tax purposes.

Another new item to the B&B study, asking whether or not the respondent provides non-financial assistance for any individuals, was designed to elaborate the additional care giving

responsibilities of the cohort. Percent agreement between interview and reinterview for this item was quite high at 88 percent. However, the relational statistic was low at 0.33. Ninety percent of respondents initially reported that they did not provide nonfinancial assistance, and of those, there was 93 percent agreement between the interview and reinterview. However, among the remaining 10 percent who initially reported that they did provide nonfinancial assistance, only 41 percent gave the same response during the reinterview. Furthermore, the relational statistic for CATI responses (0.45) was higher than it was for web respondents (0.23; t=2.14, p<0.05). This suggests that interviewers clarified the intent of this question for respondents. Revised question wording and help text should alleviate the mode difference in the full-scale study.

Table 25. B&B:93/03 reliability indices for finance items: Number of dependents and provision of nonfinancial assistance

Item series	Number of cases[1]	Percent agreement[2]	Relational statistic[3]
Number dependents claimed on 2001 taxes	99	68.7	0.79
Provide nonfinancial assistance	175	88.0	0.33[4, 5]

[1]Analyses were conducted only for respondents with determinate responses on both the initial interview and the reinterview; not all questions were applicable to all respondents.
[2]Percentage agreement is based on an exact match for nominal and ordinal measures. For continuous variables, values had to be within one standard deviation.
[3]Unless otherwise noted, relational statistic used was Kendall's Tau, τ_b.
[4]Relational statistic used was Cramer's V.
[5]The relational statistic is deceptively deflated due to insufficient variation across valid response categories. As a result, minor changes on the distribution of responses between the original and reinterview significantly lower the correlation coefficient.
SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03).

C. 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 with a nonresponse bias analysis and through quality assurance monitoring. Results for both evaluations are presented below, along with a summary of quality circle meetings held throughout data collection.

1. Nonresponse Bias Analysis

To determine whether there were any differences between respondents and nonrespondents in outcomes, a nonresponse bias analysis was conducted. Table 26 shows the results of the comparison between the B&B:93/03 field test respondents and nonrespondents on four characteristics: age, race, gender, and geographic region of last known address. No statistically significant differences were found between the distribution of respondents and nonrespondents on age, gender, or geographic region. Among the race options, a higher percentage of respondents (90 percent) were White than nonrespondents (83 percent, χ^2=9.29, p<0.005). However, no other race category yielded a statistically significant difference between respondents and nonrespondents, perhaps due to the relatively small sample sizes.

Table 26. Comparison of B&B:93/03 respondents and nonrespondents

	Respondents		Nonrespondents	
	Sample size	Percent estimate	Sample size	Percent estimate
Age				
30 to 32	299	44.4	118	46.8
33 to 35	134	19.9	58	23.0
36 to 38	36	5.3	15	6.0
39 to 41	31	4.6	8	3.2
42 or older	77	11.4	20	7.9
Missing	96	14.3	33	13.1
Race				
White	607	90.2	209	82.9
Black or African American	30	4.5	19	7.5
American Indian or Alaska Native	1	0.2	2	0.8
Asian or Pacific Islander	11	1.6	7	2.8
Other	6	0.9	1	0.4
Missing	18	2.7	14	5.6
Gender				
Male	252	37.4	107	42.5
Female	421	62.6	145	57.5
Geographic region				
New England	44	6.5	13	5.2
Mid East	98	14.6	47	18.7
Great Lakes	134	19.9	46	18.3
Plains	55	8.2	19	7.5
Southeast	194	28.8	78	31.0
Southwest	9	1.3	4	1.6
Rocky Mountains	7	1.0	5	2.0
Far West	42	6.2	24	9.5
Missing	90	13.4	16	6.4

NOTES: There were 673 respondents and 252 nonrespondents. Since these data are from a field test, they are not weighted. Details may not sum to 100 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).

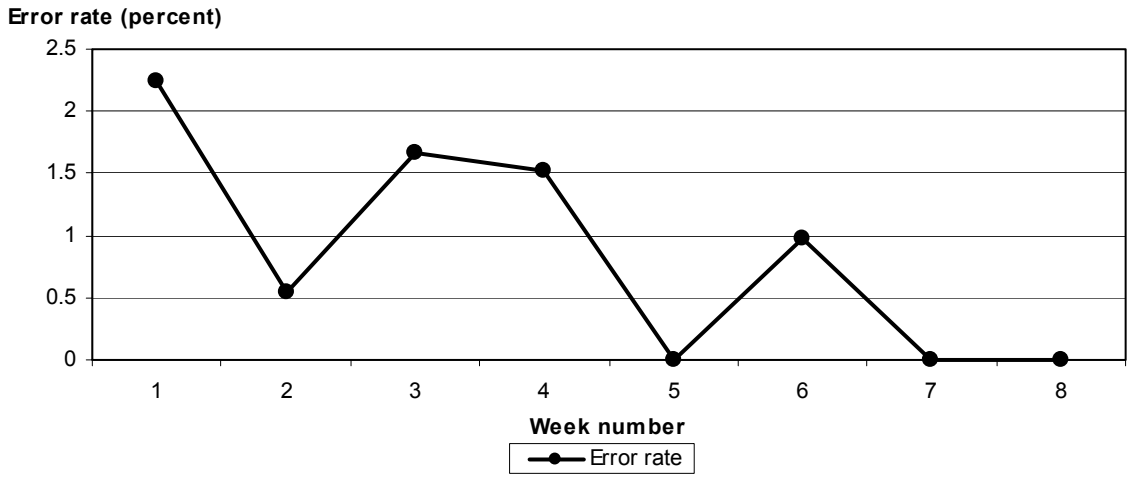
2. Quality Assurance CATI Monitoring

Regular monitoring of telephone interviews (CATI) leads to better interviewing and data quality as well as improvements in data collection costs and the efficiency of the telephone facilities. To ensure that sufficient monitoring occurred for the B&B:93/03 field test, 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 were produced. Monitoring took place during the first 8 weeks of data collection, with a total of 1,063 items monitored. After the fourth week of data collection, monitoring efforts were scaled back due to lighter caseloads. Among the 1,063 items observed,

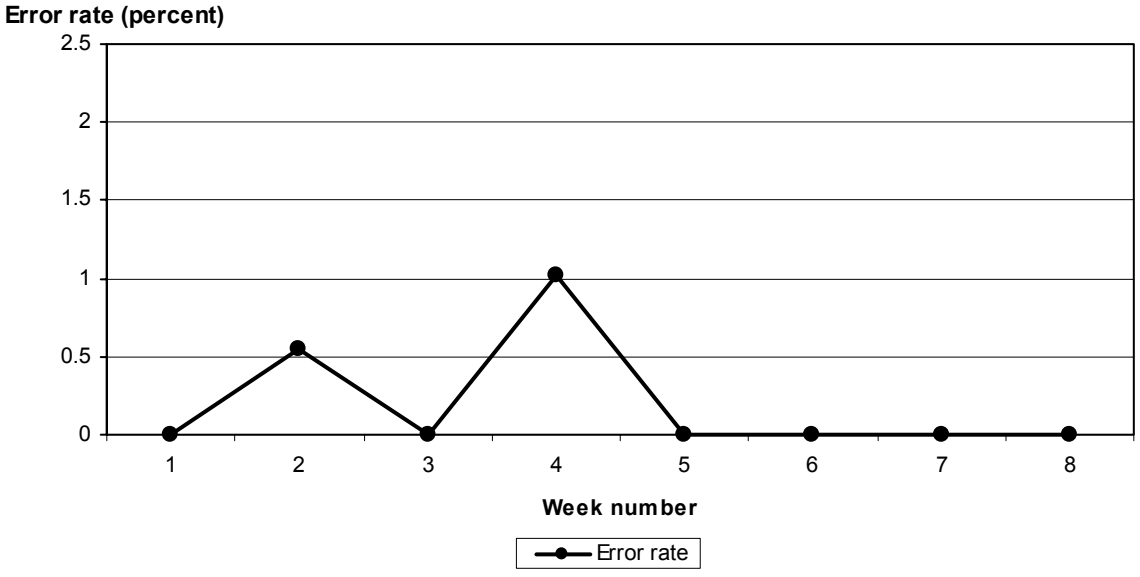
only 13 delivery errors and 3 data entry errors were observed. Error rates in delivery and data entry over the 8 weeks monitoring occurred are shown in figures 4 and 5, respectively.

Figure 4. Monitoring error rates for computer-assisted telephone interview (CATI) question delivery



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

Figure 5. Monitoring error rates for computer-assisted telephone interview (CATI) data entry



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

3. CATI Quality Circle Meetings

Quality circle meetings provided an opportunity for B&B:93/03 interviewers to discuss data collection issues with project staff. Topics addressed during these meetings included Help Desk problems, tracing and locating of respondents, and conducting efficient and effective interviews. 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.

Some of the issues covered in quality circle meetings included the following:

- Changes to the instrument: Any modifications made to the instrument once data collection began were reviewed with interviewers.

- Instrument logic: Questions regarding instrument logic were raised and discussed, and an opportunity was provided to identify modifications to the instrument and/or documentation for full-scale data collection.

- Item wording: Recommendations were discussed for clarifying question wording for both the field test and full-scale interviews.

- Help screens: Recommendations were made for modifying help text, including adding definitions, instructions, and question clarifications to assist interviewers in coding.

- Momentary web interruptions: Procedures on how to document and resolve problems with Internet connections were reviewed.

- Coding: Coding strategies were reviewed, as needed.

- Problem sheets: Telephone interviewers used electronic problem sheets to report data collection problems. Problem sheets were reviewed daily. Issues were handled immediately when necessary, then discussed with the entire group during quality circle meetings.

D. Conclusion

The primary goal of this chapter was to evaluate the quality of the data collected by the B&B instrument. Evaluations of the usability of the instrument showed that the B&B interview is easy to navigate and in general very user friendly. Two modifications will be made to the full-scale web instrument based on evaluation of the usability of the instrument. The first modification will be made to the progress bar so that progress is displayed within a section and across all sections simultaneously. The second modification will be made to the occupation coding system. Results showed that it was difficult for users to verify that the correct code had been chosen.

Data quality was also evaluated based on the effectiveness of the instrument as a multimode interview. Indeterminate rates were found to be greater for web, self-administered respondents than for CATI and CAPI, interviewer-administered respondents. The full-scale instrument will be modified to encourage respondents to provide valid responses. The reliability reinterview and the analysis of item-level mode effects uncovered similar problems on questions asking respondents to evaluate aspects of their undergraduate and graduate educations. Unreliable items will be either completely eliminated from the full-scale instrument or modified to eliminate mode effects.

The last major area that was employed to measure data quality was the effectiveness of the data collection design. No major data quality issues were uncovered based on the nonresponse bias analysis and quality assurance CATI monitoring.

Chapter 5
Recommendations for the Full-Scale Study

The purpose of the B&B:93/03 field test was to test procedures and inform planning for the full-scale study. Essential aspects of the field test survey design and instrumentation, including the design and implementation of a single web-based instrument for self, telephone, and in-person interviewing, were conducted successfully. Chapters 3 and 4 of this report documented those field test outcomes and evaluation results that warranted procedural and/or substantive modifications to the full-scale study design. Recommended changes to the sampling design, tracing and data collection plans, and instrument are summarized below.

A. Sampling Design

For the full-scale study, the sample will consist of all B&B:93/97 respondents, plus approximately one-third of nonrespondents. To select the nonrespondent subsample, all B&B:93/97 nonrespondents will be stratified by advance (batch) tracing outcome and response status for the base year (NPSAS:93) and first follow-up (B&B:93/94) interviews. Those B&B:93/97 nonrespondents most likely to be located and interviewed will be oversampled.

Response rates obtained in the B&B field test will provide useful information for determining the nonrespondent sample allocation for the full-scale study. Table 27 provides the B&B:93/03 field test sample sizes and response rates by advance tracing outcome for all sample members, and separately for B&B:93/97 respondents and nonrespondents. These rates, along with advance tracing outcomes for the full-scale sample and a review of response rates from comparable studies, will be used to identify several possible sample allocations and associated design effects in order to determine the optimal sample allocation for full-scale data collection.

Table 27. B&B:93/03 field test sample sizes and response rates, by B&B:93/97 field test advance tracing outcome and response status

Tracing outcome and prior response status	B&B:93/03 field test sample size	B&B:93/03 field test respondents	Response rate (percent)
Total	925	673	72.8
Located during advance tracing	879	647	73.6
Not located during advance tracing	46	26	56.5
B&B:93/97 field test respondents	850	637	74.9
Located during advance tracing	808	612	75.7
Not located during advance tracing	42	25	59.5
B&B:93/97 field test nonrespondents	75	36	48.0
Located during advance tracing	71	35	49.3
Not located during advance tracing	4	1	25.0

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

B. Tracing and Locating

Overall, the tracing and locating systems customized for the B&B:93/03 field test worked well, efficiently handling the extensive locating information available for each sample member. Since the web self-administered option resulted in a higher response rate than initially expected, it will be important that the full-scale data collection mailings reach the sample member, thus maximizing the potential for self-administered interviewing. The initial address update mailing will be addressed to both the sample member and the parents of the sample member to increase the likelihood that sample members will receive it quickly. Given the longer data collection period for the full-scale study, there will be additional time available to process all address updates from the early mailings, including the lead letter mailing.

C. Interviewer Training

Telephone data collection staff gave favorable reviews about project training. Among the strengths noted were the enthusiasm of the project training team, an increased emphasis on how to answer respondent questions, availability of "cheat sheets" for using the coding systems, and a training schedule that allowed time for more individual practice. Some aspects of training will be modified for the full-scale study in response to interviewers' suggestions for improving the training process. These include developing training examples from actual field test data when preparing the full-scale training materials and simplifying the question look-up utility.

D. Help Desk

According to Help Desk staff, sample members particularly liked the freedom and convenience offered by the self-administered web interview. Security of the data was mentioned as a concern by only a small number of sample members. Help Desk staff felt adequately prepared to handle most of the technical problems encountered by respondents and reported that the system used to document the calls worked well.

Help Desk staff offered several recommendations for improving operations for the full-scale study, most of which concerned the mechanics of the new Help Desk application. These

issues will be addressed in preparation for the full-scale study. One additional problem was that, since the calls to the Help Desk were infrequent, staff tended to forget the details of handling various types of problems presented during training. While a higher number of calls during full-scale data collection is anticipated, project staff will incorporate a review of potential Help Desk problems during regular quality circle meetings.

E. Early Response Incentive

The early response incentive experiment described in chapter 3 compared response rates for two groups—those who received an offer of a $20 cash incentive for completing the web self-administered interview within 10 days of the start of data collection and those who did not receive the incentive offer. Although the number of observations was small, the results suggested that the incentive offer did have a positive effect on early response. More sample members who received the incentive responded within the 10-day response period compared to those who did not receive the offer. Because a high rate of web response decreases the need for tracing and more expensive telephone and field interviewing, overall data collection costs can be reduced. Therefore, the early response incentive will be offered to all sample members who respond by web within the first 3 weeks of the start of data collection.

F. Field Interviewing

Overall, field interviewers reported receiving good tracing information for field locating. Field interviewers found that a number of "bad" telephone numbers provided by the Telephone and Internet Operations (TIO) unit were incorrect only because the area codes were wrong. Tracing staff will be asked to include a step to check area codes during tracing activities for the full-scale study.

G. Web Screen Design

Very few modifications will be made to the interview screen design. The header art will change slightly to be consistent with the newly designed home page for the study web site. Additionally, the progress bar will be the only respondent motivator in the full-scale study, since results of the field test experiment comparing the effectiveness of a progress bar to motivational statements between sections were inconclusive. The bar itself will be modified from the field test, however. Rather than displaying progress solely within the current section, the bar will show progress within each section and across the five sections of the interview simultaneously.

H. Instrumentation

Revisions will be made to the field test interview based on examination of the field test results presented in chapters 3 and 4 and discussions with the technical review panel (TRP; a list of TRP members and their affiliation is provided in appendix D). The most salient modifications to the instrument are described below, and modifications to the data elements are shown in table 28. The set of data elements used for the field test instrument is included in appendix E.

Given the differences in interview times across modes, the goal will be to develop a full-scale telephone interview that averages 25 minutes in length. Average length of the Web interview will vary depending on connection speed, and sample members with slower connection

speeds will be encouraged to call the telephone unit to complete a telephone interview, rather than attempting the web interview.

During the field test, an unusually high rate of indeterminate responses—both "don't know" and refusal responses—was observed in the interview section on finances. Compared to telephone and field interview rates, self-administered indeterminate rates for the finance questions were markedly higher. The availability of explicit "don't know" and "decline to answer" options on each self-administered interview screen may increase the likelihood that respondents use them.

For the full-scale interview, the explicit indeterminate options will be removed from each interview screen. An effort will be made to convert indeterminate responses only for critical items in the full-scale interview. That is, should a respondent click the continue button without providing a response, he/she will be provided with a brief explanation of the importance of the data requested and offered the question again. For income and other financial questions identified as critical, categorical response options rather than explicit response requirements will be provided.

To improve the responses obtained in the major and occupation coding systems, on-screen descriptions comparable to those available for the industry coder will be provided to better define the response categories provided. Help text and other on-screen text will also be reviewed and improved as necessary to provide additional support to respondents and interviewers. During full-scale data collection, a random sample of 10 percent of coding responses will be reviewed regularly so that any remaining problems can be addressed with interviewers early in data collection.

To ensure that responses are not influenced by the order in which the items are presented, opinion questions with multiple items will use a randomly generated order of presentation. In addition, as noted in table 28, those field test items that require a rating of "very," "somewhat," or "not important" will be redesigned to require that respondents select only those aspects considered "very important." Such a change is expected to save time in the interview, since the questions will no longer require an explicit response for each item.

Table 28. B&B:93/03 field test data elements recommended for modification for the full-scale instrument

Data element	Field test outcome	Modification
Other postbaccalaureate education		
Since 1997, whether enrolled in any vocational or technical diploma or bachelor's degree programs	Previous level of detail for undergraduate degree program required too much recall and yielded only a small number of cases who had pursued an additional undergraduate degree program. Previous level of detail for more than one license/certification also required too much recall and did not apply to a sufficient number of cases to be analyzable.	For each undergraduate degree program, will ask only degree attempted and completed. For license/certification, will ask whether any new license/certification was earned since the last interview and, if yes, whether it was a renewal or upgrade of an existing license/certification. If not, will ask all items for "most recent" occurrence, and what was attempted and completed for all.
Month and year of first/last enrollment in the program	Recalling months was burdensome and information about the length of time taken to complete the program (or completed so far) can be asked directly with more certainty rather than calculated from potentially unreliable dates.	For the most recent license/certification, will ask whether currently enrolled, and collect amount of time spent working on the license/certification, rather than start and end dates.
Whether completed diploma or degree program		For license/certification, collect whether completed or date expected.
What aspect(s) of undergraduate education stands out as influential or important (instruction received, major, extracurricular activities, etc.)	Some items were shown to have poor temporal stability.	Items with poor temporal stability have been deleted. Will collect those aspects of the undergraduate education which stand out as *very important* to their lives now; aspects of undergraduate education include major, liberal arts courses, professional courses, quality of instruction, internship, and other work while enrolled.
How would respondent evaluate undergraduate education with respect to relationship to work, preparation for life, price, social contacts, health, financial security, overall happiness	Some items were shown to have poor temporal stability.	Items with poor temporal stability have been deleted. Response format will focus on rating of "very important." Will ask respondents to rate the undergraduate education with respect to work and career, further education, and financial security, and evaluate whether the undergraduate education was worth the cost, time, and effort required.

Continued

Table 28. B&B:93/03 field test data elements recommended for modification for the full-scale instrument—Continued

Data element	Field test outcome	Modification
Other postbaccalaureate education (continued)		
(If completed a graduate degree) What aspect(s) of graduate education stands out as influential or important (instruction received, major, extracurricular activities, etc.)	Some items were shown to have poor temporal stability.	Items with poor temporal stability have been deleted. Response format will focus on rating of "very important." Will ask respondents to rate which aspects of the graduate education are "very important"t to their lives now. Aspects of the graduate education include field of study, quality of instruction, internship and other work while enrolled.
(If completed a graduate degree) How would respondent evaluate graduate education with respect to relationship to work, preparation for life, price, social contacts, health, financial security, overall happiness		Unreliable items have been deleted. Response format will focus on rating of "very important." Will ask respondents to rate the graduate education with respect to relationship to work, career, and financial security; and evaluate whether the graduate education was worth the cost, time, and effort required.
Employment		
Since 1997, whether ever not employed (unemployed or out of the labor force) for a period of at least 3 months	Unemployment questions collected by spells were time-consuming. Dates were difficult to recall.	Respondent will be asked for summary information about the number of spells of unemployment and out of the labor force. Total amount of time, length of longest spell, and reasons for being out of the labor force will be assessed.
If unemployed, whether received unemployment compensation		Will ask for summary measure rather than spell-specific measure.
Status as of April 2003		Timing of data collection requires reference date to be February rather than April.
Job-related information for April 2003 or most recent job (if more than one job, information for primary employer)	The fixed reference month produced recall problems and confusion about the targeted job.	Job-related information will be asked only for the current or most recent job.
How important were undergraduate education, graduate education, on the job experience, other education experience, and other work experience for that job		The item about work experience on the current job has been deleted. Response format will focus on rating of "very important."
Information about those not employed in April 2003		The reference month will be changed to February 2003.

Continued

Table 28. B&B:93/03 field test data elements recommended for modification for the full-scale instrument—Continued

Data element	Field test outcome	Modification
Teachers		
Whether working toward/completed national teacher certification (NBPTS)		This is a new form of certification that many states are encouraging through financial incentives. Item will be added.
Whether undergone professional development in the use of technology in the classroom		Related to policy initiatives regarding the use of technology. Item will be added.
Number of students for whom responsible in most recent semester of the job		Item will be added to assess workload.
Whether taught special student populations (e.g., AP/honors, limited English proficient)		Item will be added to assess workload.
Factors that make you want to stay in teaching		Specific reasons provided will be edited to ensure comparability with other data sources and current policy initiatives. Open-ended option will be removed.
Factors that make you want to leave teaching		Specific reasons provided will be edited to ensure comparability with other data sources and current policy initiatives. Open-ended option will be removed.
Finances and debt		
Other nonwage income of respondent or spouse/partner		Total household income from all sources will be asked directly rather than calculated from other reported amounts.
If education loans completely repaid, when finished	The month of repayment was difficult to recall.	The item will be modified to ask for year only.
When payments started	The month of repayment was difficult to recall.	The item will be modified to ask for year only.
What year spouse began repayment of education loans		Will be added to provide comparable information to items about respondent's own loans.
What year spouse ended repayment (if no longer in repayment)		Will be added to provide comparable information to items about respondent's own loans.
Living arrangement: own/rent/neither		Will be added as gate question for mortgage/rent expenses.
Net worth	Individual items showed high rate of indeterminate responses.	Rather than asking for dollar values of various assets and debts, will just collect whether they are present.
Total number of dependents claimed when filing income taxes	Item is redundant.	Item will be deleted.
Number and relationship to respondent of household members	Item is redundant.	Item will be deleted.

Continued

Table 28. B&B:93/03 field test data elements recommended for modification for the full-scale instrument—Continued

Data element	Field test outcome	Modification
Family formation and civic participation		
(If children under 18) Hours per week of nonparental child care	Hours per week were difficult to estimate.	Will ask whether preschool children are enrolled in full-time or part-time child care and whether school-age children are enrolled in before- or after-school care
(If school-age children) Whether in public/private/home schooling		Will be added to provide a picture of cross-generational education patterns.
Number of individuals for whom have financial responsibilities (e.g., spouse, children, parents, other family members, or others)	Item is redundant.	Item will be deleted.

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

I. Conclusion

The purpose of the B&B:93/03 field test was to test fully all data collection procedures in preparation for the full-scale study. Especially important to the success of full-scale data collection was the successful design and implementation of a multimode, web-based instrument for administration by both sample members and interviewers. As discussed in this report, the B&B:93/03 instrument was effective for self-administration and as a telephone and field interview, and therefore will require only minor modifications of its design for full-scale administration. Question wording will be modified in response to recommendations from the Technical Review Panel. Because the web is still a relatively new data collection technology, the design strategies implemented for the full-scale study will be reevaluated following its completion to further explicate the optimal design for web-based data collection.

The tracing and locating procedures implemented for the field test, as well as the Help Desk support provided to web users, were successful for the field test and will be employed again as designed for the full-scale study. In addition, an incentive will be offered to sample members who complete a self-administered interview on the web within the first 3 weeks of data collection. Since the field test was on an abbreviated schedule, the increased timeframe for the early response incentive should increase web response rates above the 26 percent observed in the field test.

References

Agresti, A. (1984). *Analysis of Ordinal Categorical Data*. New York, NY: Wiley & Sons.

Cannell, C.C., Miller, P.V., and Oksenberg, L. (1981). Research on Interviewing Techniques. In S. Leinhardt (Ed.), *Sociological Methodology* (pp. 389-437). San Francisco: Jossey-Bass.

Cannell, C.F., Oksenberg, L., and Converse, J.M. (1979). *Experiments in Interviewing Techniques*. Ann Arbor, MI: Institute for Social Research, University of Michigan.

Cohen, J. (1987). *Statistical Power Analysis for the Behavioral Sciences* (rev. ed.). Hillsday, NJ: Erlbaum.

Dillman, D.A. (2000). *Mail and Internet Surveys: The Tailored Design Method*. New York, NY: Wiley.

Duffer, A., Lessler, J., Weeks, M., and Mosher, W. (1994, August). *Effects of Incentive Payment on Response Rates and Field Costs in a Pretest of a National CAPI Survey*. Paper presented at the annual meeting of the American Statistical Association, Toronto, Canada.

Kendall, M. (1945). The Treatment of Ties in Rank Problems. *Biometrika*, *33*: 81-93.

This page is intentionally left blank.