

Challenges of Designing and Implementing Multimode Instruments

**Jennifer Wine, Melissa Cominole, Ruth Heuer,
and John Riccobono
RTI International**

Over the last 10 years, nationally-representative, cross-sectional and longitudinal telephone surveys of postsecondary populations, conducted for the National Center for Education Statistics (NCES), have evidenced the same decline in response rates observed in other populations. Locating and interviewing such a highly mobile population has become increasingly difficult and costly.

In 2002, RTI International (RTI) conducted the first NCES postsecondary student survey that offered a web, self-administered option as part of a multimode data collection effort. By not requiring extensive interviewer tracing and interviewing, web, self-administration offered a cost-effective alternative for surveying such a technologically-savvy group. Since 2002, RTI has conducted eight studies of postsecondary students and faculty that combine web, self-administered and telephone survey methodologies:

- The 2002 field test and 2003 full-scale implementations of the 1993/2003 Baccalaureate and Beyond Longitudinal Study (B&B FT and B&B FS);
- The 2003 field test and 2004 full-scale implementations of the 2004 National Postsecondary Student Aid Study (NPSAS FT and NPSAS FS);
- The 2003 field test and 2004 full-scale implementations of the 2004 National Study of Postsecondary Faculty (NSOPF FT and NSOPF FS);
- The 2004 field test of the 2004/06 Beginning Postsecondary Students Longitudinal Study (BPS FT); and
- The 2004 field test of the 2000/06 Education Longitudinal Study (not discussed here).

Additional information on these studies can be found on the NCES web site at <http://www.nces.ed.gov/surveys/>.

Introduction of a self-administered option for these studies has helped to achieve the targeted response rates and ensure population coverage, and its flexibility has seemed particularly suited to postsecondary student and faculty schedules. However, by moving to a multimode approach, we were unwilling to lose particular aspects of the telephone interview experience, specifically, the advantages of the interviewer's presence for response conversion, clarification of question text, resolution of inconsistencies, and coding of responses. Since our first experience in 2002, therefore, we have been working to blend the best features of the telephone interview with those of the web, self-administered interview to develop a single instrument for use in multiple modes. Throughout the evolution of the instrument in these studies, we continue to work to minimize any differences in respondent behavior due to mode while maintaining high data quality.

Telephone Prompting

Faced with declining response rates, the option to complete a web, self-administered interview rather than a telephone (CATI) interview offered a cost effective alternative. At the same time, the savings gained by not paying for extensive interviewer involvement would allow the introduction of incentives to encourage web response prior to the start of CATI data collection. In the B&B FT in 2002, we found that offering incentives for completion of a self-administered, web interview prior to the start of telephone interviewing did increase web completion rates over those when no incentive was offered for an early web complete. Thirteen percent of incentivized sample members completed the web, self-administered interview in the early response period compared to only 9 percent who were not offered an incentive ($z=1.9$; $p<.05$).

When the early response incentive was offered to all sample members in the B&B FS study a year later, we obtained 18 percent of completes during the early web response period. That rate climbed another 10 percentage points in 2004 when 28 percent of overall NPSAS FS completes were obtained during the early web response period. Among NSOPF FS faculty surveyed in 2004, 44 percent completed.

During the early response period, which typically covers the first 4 weeks of data collection, specially-trained telephone interviewers monitor the Help Desk, but no outgoing telephone interviewing occurred. If sample members do not complete an interview immediately upon initial contact, the challenge was in keeping them engaged until the interview was completed. Postcards and e-mailed messages to sample members had limited effectiveness. Could telephone interviewers, already monitoring the Help Desk, encourage increased participation among procrastinators?

In the BPS FT in 2005, postcards and e-mails were supplemented with telephone prompts initiated by Help Desk staff. Sample members were prompted with a scripted message either directly on the telephone, or indirectly through messages left with household members or on an answering machine reminding them of the impending deadline for receipt of the early response incentive. A comparison of response rates showed that prompted sample members were twice as likely to complete the web, self-administered interview during the early response period than were non-prompted sample members (21 percent compared to 9 percent; $p<.01$). By the end of the response period, almost 39 percent of interviews had been completed, and, as shown in figure 1, average call counts to obtain a web complete (6.5 calls) still remained significantly lower than the count required to complete a telephone interview (14.9 calls; $t=-8.58$; $p<.0001$).

Interview Burden

Ensuring that respondents in a multimode data collection do not experience excessive burden (measured as total time in the interview) for selecting one response mode over another has been a formidable challenge. Time in the interview differed significantly by mode in the B&B FT. Web, self-administered respondents averaged 40.5 minutes for a completed interview, compared to 35.7 for the interviewer-administered interview ($t=3.51$, $p<.001$). When broken into its component parts, however, the

interview itself – that is, the time required for questions to be read or administered and a response made – did not differ between modes. The 5 minutes' difference in burden was due almost entirely to the differences in transit time, the time required to transmit data from the respondent to the server and back to the respondent. Whereas transit time averaged 4.4 minutes, web, self-administered respondents averaged 12 minutes ($t=13.15$; $p<.0001$).

By recommending to sample members that they only attempt the web, self-administered interview using a high speed connection, burden associated with completing the same interview in the B&B FS did not differ for web, self-administered (34.4 minutes) and telephone (34.8 minutes) modes. While the overall experience of the interview burden was the same, significant differences were observed in the components. Time in the interview was shorter for web respondents (24.4 minutes) than for telephone respondents (29.1 minutes; $t=-23.05$; $p<.0001$) while transit times for a web interview averaged about twice those for telephone (10 minutes compared to 5.7 minutes; $t=34.7$; $p<.0001$). Fortunately for respondents, the overall perception of interview burden was the same irrespective of mode.

In 2004 again, the NPSAS FS web, self-administered and telephone interviews showed the same pattern. Overall average time in the interview was approximately the same, differing by less than one minute (27.9 minutes for web, 27.0 minutes for telephone; $t=9.48$; $p<.0001$). However, component times differed considerably by mode. On-screen times were significantly less for the self-administered interview (22 minutes) than for telephone (24 minutes). Transit times for web interviews averaged almost twice that of telephone interviews (6.0 compared to 3.1 minutes; $t=97.06$; $p<.0001$).

It would seem then that it is possible to achieve comparable average interview times across modes in a multimode data collection. On screen times tend to be higher among telephone interviews, most likely because respondents and interviewers engage in conversation, and, depending on the type of question, interviewers must read all response options. Transit times are consistently higher when the interview is self-administered. Debriefings of web respondents have shown that approximately 10 percent of web interviews are being completed using dial-up modems which, for the NPSAS FS, required almost 3 times the transit time of fast connections (14 minutes compared to 5 minutes; $t=-85.4$; $p<.0001$). All telephone interviewers access the survey using high speed, T1 lines.

Help Desk

In order to provide assistance to web, self-administered respondents like that provided respondents by interviewers over the telephone, a Help Desk remained available throughout data collection for each multimode study. Help Desk staff were trained to handle any sort of problem that might be encountered by sample members attempted to completed the interview over the web. A software program recorded each Help Desk incident that occurred.

Consistently across student samples, the most common Help Desk problem involved missing Study IDs and/or passwords (50 percent of problems encountered in the BPS FT). Another 20 percent of calls were related to browser settings and computer problems. Questions about the study and instrument content accounted for only about 3 percent of problems. In contrast, among NSOPF FS faculty, calls to the Help Desk focused on questions about the study (19 percent).

Item Reliability

In the field tests for each of the studies, reliability reinterviews, typically containing items that were new or had been troublesome in past administrations, were administered to a subset of respondents to the main interview. With the transition to multimode data collection, items were evaluated for temporal stability, for all respondents and also by administration mode. The reinterview was conducted in the same mode as the original interview.

For example, the reinterview sample for the B&B FT was evenly divided by mode of response with 250 self-administered respondents and 250 telephone interview respondents chosen. Key information from the initial interview was preloaded for the reinterview to ensure that questions were asked in the same way and with the same wording across the two interviews. Reinterviews were conducted in the same mode as the initial interview, about 3 weeks following its completion. By the end of data collection, 36 percent of the self-administered respondents and 75 percent of the telephone interview respondents completed the interview.

Responses in the initial interview and the reinterview were compared to determine the percentage of reinterview responses that matched the original responses from the main interview. For categorical data, agreement required an exact match of interview and reinterview responses. For continuous data, responses were considered to agree when reinterview values were within one standard deviation of the main interview values.

As part of the reinterview analysis, percent agreement rates for self-administered and telephone interview respondents were compared to determine the extent of difference in rates due to mode of interview completion. Only two items in the reinterview—"value of undergraduate internship and other work experience" (81 percent web compared to 71 percent telephone; $\chi^2 = 4.25$; $p < 0.05$) and "importance of undergraduate education to current employment" (70 percent web compared to 80 percent telephone; $\chi^2 = 4.33$; $p < 0.05$)—showed statistically significant differences by mode. No other differences by mode were observed.

Although telephone respondents have tended to participate at a higher rate than web, self-administered respondents, no consistent mode differences in reliability have been observed across studies. In the few cases where statistical differences in item reliability were observed, there was no pattern indicating better reliability in one mode or the other. When mode differences were observed, very few items (e.g, 2-5 items from the reinterview) were affected. No mode differences in item reliability were found in either the NSOPF FT in 2004 or the BPS FT in 2005.

Coding Systems

Coding systems have been used routinely in telephone interviewing to categorize information text strings. The coding systems used for the NCES studies, which are most often used to code major/field of study, occupation, and industry text strings, have evolved over time. Most recently, they have been redesigned to maximize ease of use for both interviewers and self-administered respondents in multimode surveys.

A recent experiment conducted as part of BPS FT compared the error rates among two different coding systems overall and by administration mode. The effectiveness of the procedures used to code *major/field of study* and *occupation* were assessed by expert coders who evaluated a random sample of major and occupation codes representing 50 percent of text strings provided by self-administered web and CATI/CAPI respondents.

Two coding systems were used for categorizing occupation, O*NET-A and O*NET-B.¹ Two expert coders assessed each case. Per respondent, one coder used the same coding method employed in the original interview, while the other used the alternative coding method. Percent agreement was calculated for each possible combination of variable type (*major* or *occupation*), coding method (double drop-down or assisted for *major*, and O*NET-A or O*NET-B for *occupation*), and survey mode (self- or interviewer-administered).

Percent agreement for each comparison is presented in table 1. Expert recoding of the *major/field of study* assisted coding system did not show differences across modes (self-administered interview versus CATI/CAPI), indicating that both coding systems worked equally well with both interviewers and self-administered respondents. Demonstrating the reliability of the coding methods and their corresponding administration modes, percent agreement did not change significantly when the expert coder used the opposite coding method employed in the interview. Likewise, the *occupation* coding systems tested (O*NET-A and O*NET-B) failed to show statistically significant mode differences. Self-administered respondents and trained interviewers are equally successful at using the online coding systems.

Item Nonresponse

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

To minimize item-level nonresponse, several changes were made to the full-scale instrument. First, the “don’t know” and “refuse” options were removed from the screen

¹ Developed for the U.S. Department of Labor’s Employment and Training administration, O*NET is a comprehensive database of worker attributes and job characteristics. For more information, visit <http://online.onetcenter.org/>.

entirely. Instead, respondents could use the “continue” button to proceed without answering if an answer was unknown or they refused to answer a particular item. Second, if respondents continued through three consecutive items without providing a response, a generic pop-up box was presented to remind them of the importance of their continued participation in the interview. The pop-up box is presented in figure 2.

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

With the presentation of the conversion text, a “don’t know” option was added to the choices of response options. For the four key items on income, the conversion text was presented but, rather than providing a specific dollar value, respondents selected from among several categorical income ranges, and the “don’t know” option. A sample of one of the items with conversion text is shown in figure 3. Once presented with the conversion text, respondents could select one of the original response options, choose the “don’t know” option, or continue without providing a response. The effectiveness of this approach to converting indeterminate responses is discussed below.

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

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

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

In the teaching section, teachers were asked about the dates they began and ended teaching jobs held since 1997 and about the schools in which they taught during the 6 years elapsed between 1997 and 2003. Up to 16 percent of teachers asked could not provide information on the school that employed them prior to their current school.

However, since the questions required recall of dates and other details, the observed nonresponse was likely the result of not knowing an answer rather than refusing to provide an answer. Twelve percent of teachers did not provide a response for the item asking for the primary reason they left teaching. No differences were observed in the rates of indeterminacy by mode for the items in the teacher section.

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

As shown in table 2, only one item in the background section, which asked respondents who had previously indicated having a disability to report their disabling condition, had an indeterminacy rate of at least 10 percent. In fact, the rate of nonresponse (31 percent) was quite high, perhaps due to the sensitivity of the question. No mode differences were observed in the rates of indeterminacy for this item.

Table 3 presents the results of offering conversion text for 20 key items in the full-scale interview. For each item targeted, the number reaching the text is shown, together with the number and percentage providing a definitive response (i.e., either selecting a response option or choosing "don't know"). Also shown is the number and percentage who continued through the item without providing a response, effectively refusing to respond. The conversion text successfully converted responses from missing for 18 of the 20 items targeted. The item with the lowest conversion rate at 33 percent—currently looking for work—had fewer than 10 nonrespondents. After reading the conversion text, 52 percent of respondents who initially did not provide a response to the monthly rent or mortgage payment item ended up providing a response (either definitive or don't know).

Responses could have been converted to one of the provided response options or to a "don't know" response. As shown in table 4, the percentage of respondents providing an explicit response (i.e., a response other than "don't know" for an item) ranged from 0 percent to 100 percent. Even for those income items considered most sensitive, response conversion was high.

Help Text

Help text was made available for every screen in all study instruments. This information was considered useful for self-administered respondents because it provided

detailed information on the intent of the item, clarification of response options, and some examples. Help text provided interviewers with quick access to additional information for use in clarifying questions. Item-level counters tracked the number of times each help text screen was accessed, making it possible to identify screens that were confusing to interviewers or respondents.

The screen-level rate of help text access was typically below 2 percent for most of the screens in the studies reviewed. Help text access rates were analyzed overall and by administration mode. Findings on help text usage have been consistent across studies that have used the multimode web-based data collection strategy. Results from the NPSAS FS, for example, showed that across all interview forms, interviewers accessed help text at a higher rate than did self-administered cases (1.3 percent compared with 0.3 percent, respectively; $t = 6.43$, $p < 0.0001$).

Conclusion

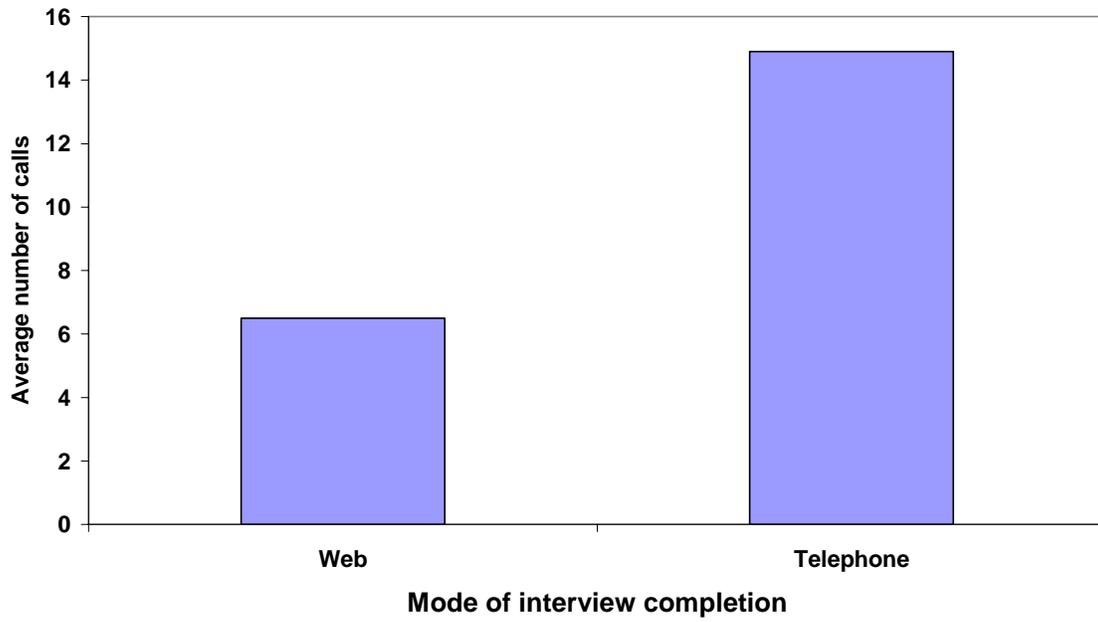
A primary goal of our survey work has been to improve response rates without compromising data quality. When web surveys were first introduced, researchers were optimistic that they would reverse the trend of declining response rates. While this has not proven to be the case, we have found that using a blended mode approach that includes a web self-administered questionnaire as well as CATI and CAPI methodologies has increased response rates among our technology-savvy population of postsecondary students and faculty. The appeal of this multiple mode approach is that it provides more flexibility to sample members, allowing them to complete the questionnaire at a convenient time and in a manner that suits their interaction preferences. Catering to the preferences of sample members makes it easier for them to complete the questionnaire while at the same time reducing our costs (i.e., not having to staff a call center 24 hours per day).

The challenge for instrument developers has been to make the survey experience the same for all respondents (to the extent feasible), regardless of mode, i.e., make the modes functionally equivalent. The blended methodology approach allows survey researchers to retain the benefits of telephone interviewing (e.g. nonresponse conversion, clarification, probing) while introducing the flexibility, convenience, and cost benefits of self-administration.

Results presented earlier from timing analysis, reliability reinterviews and coding systems show little or no impact of mode effects on data quality. Item nonresponse has shown some modal differences, although we have strived to minimize these differences through the elimination of on-screen don't know and refuse options, the use of conversion text, help text, and other techniques in lieu of interviewer probing for web respondents. Item indeterminacy has declined as a result of the design changes that have been introduced.

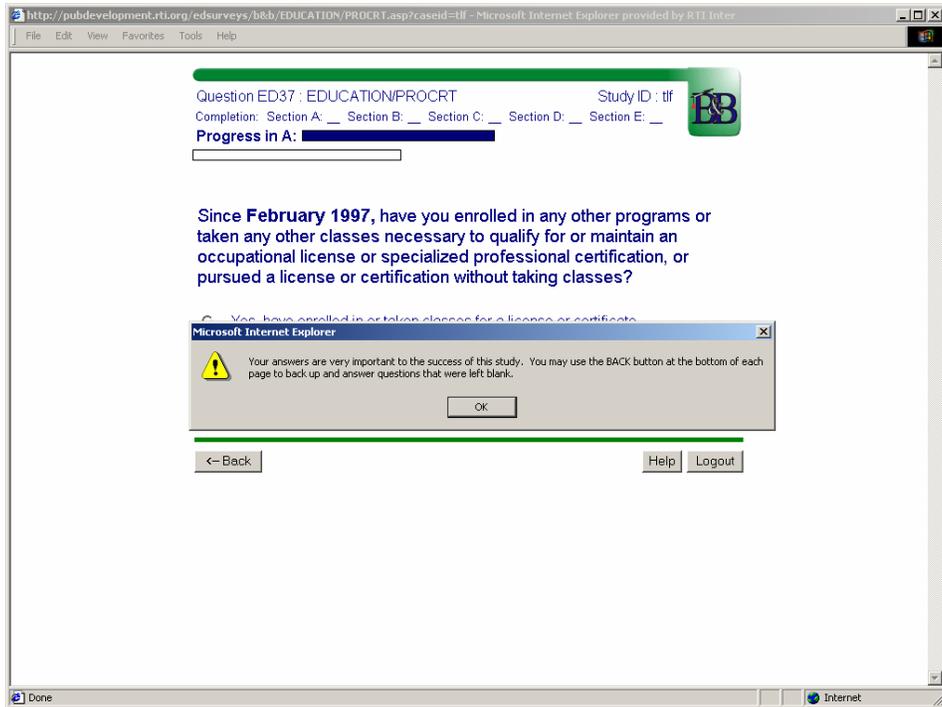
With each study conducted, new lessons are learned about the blended methodology approach: what works well and what methods are not as successful as expected, and techniques and strategies are refined and customized and tailored to the specific needs of the target population based on what experience has taught us. In the few years that the web-based multi-mode data collection has been used, the approach has changed to accommodate what we know about respondents, while achieving the goal of collecting high quality data. The lack of any substantial mode effects on data quality is evidence in support of implementing this cost-effective method for combating declining response rates while retaining superior quality data.

Figure 1. Average Call Counts to Obtain Web, Self-Administered and Telephone Interviews



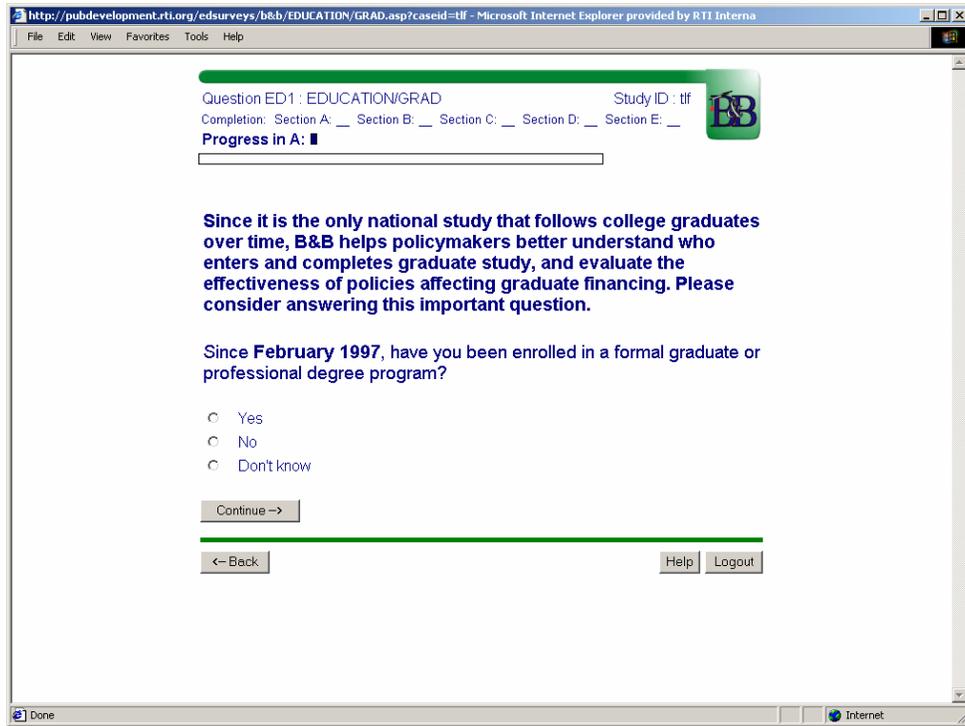
SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004/06 Beginning Postsecondary Student Longitudinal Study (BPS:04/06) Field Test.

Figure 2. Pop-up box presented when respondents failed to respond to three consecutive questions in the B&B:93/03 full-scale interview



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

Figure 3. Sample response conversion text presented when respondents to the B&B:93/03 full-scale interview did not respond to key interview items



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

Table 1. Major and occupation coding quality analysis, by percent and number in agreement with expert coders using same or different coding method

Method used by respondent	Agreement when expert coder used same method		Agreement when expert coder used different method	
	Number	Percent	Number	Percent
Major				
Double drop-down				
Self-administered interview	90	78.4	80	72.6
CATI/CAPI	70	73.6	60	74.4
Assisted coder				
Self-administered interview	90	85.3	80	76.9
CATI/CAPI	80	78.9	80	82.5
Occupation				
O*NET-A				
Self-administered interview	30	75.7	30	64.1
CATI/CAPI	50	65.2	40	61.4
O*NET-B				
Self-administered interview	20	66.7	20	64.7
CATI/CAPI	60	80.8	60	77.3

NOTE: Numbers have been rounded to the nearest 10. Detail may not sum to totals because of rounding. CATI = computer-assisted telephone interview; CAPI = computer-assisted personal interview. O*NET-A and O*NET-B are based on the U.S. Department of Labor's Employment and Training Administration's O*NET, a comprehensive database of worker attributes and job characteristics. For more information, visit <http://online.onetcenter.org/>.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004/06 Beginning Postsecondary Students Longitudinal Study (BPS:04/06) field test (forthcoming).

Table 2. Interview item nonresponse for items with more than 10 percent missing

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

¹ Item nonresponse rates were calculated based on the number of sample members for whom the item was applicable and asked.

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

Table 3. Effectiveness of directed text in converting nonresponse to key interview items: 2003

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

Rounds to zero.

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

Table 4. Effectiveness of directed text in evoking an explicit response to key interview items: 2003

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

Rounds to zero.

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