

2008/18 BACCALAUREATE AND BEYOND (B&B:08/18) FIELD TEST

Supporting Statement Part B
OMB #1850-0729 v. 11

Submitted by
National Center for Education Statistics
U.S. Department of Education

January 2017

Contents

B. Collection of Information Employing Statistical Methods.....	1
1. Respondent Universe.....	1
2. Statistical Methodology.....	1
a. Field Test Design.....	1
b. Full-scale Design.....	2
3. Methods for Maximizing Response Rates.....	2
4. Tests of Procedures and Methods.....	5
C. References.....	13

Tables

Table 1.....Distribution of the B&B:08/18 field test and full-scale sample by interview response status for NPSAS:08, B&B:08/09, and B&B:08/12, among NPSAS:08 study members ¹	2
Table 2: Items selected for the mini survey.....	8
Table 3: Items selected for the forced choice experiment.....	10
Table 4. Detectable differences (two group comparisons).....	12
Table 5: Detectable differences (stacked method, dependent samples).....	13

B. Collection of Information Employing Statistical Methods

This submission requests clearance for the 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18) field test data collection instrument and methods. B&B:08/18 is the third and final follow-up of sample members from the 2007-08 National Postsecondary Student Aid Study (NPSAS:08) who were baccalaureate recipients during the 2006-07 (field test) and 2007-08 (full-scale) academic years. For details on the NPSAS:08 sampling design and the final field test and full-scale study designs, see the Supporting Statements Part B for the NPSAS:08 Field Test and NPSAS:08 Full Scale (OMB#1850-0666), the B&B:08/09 field test and full-scale collections (OMB# 1850-0729, v. 2), and the B&B:08/12 field test and full-scale collections (1850-0729; v. 7-10). Specific plans for the B&B:08/18 field test are provided below. Note that overlapping B&B cohorts are in different stages of data collection. A newer cohort of baccalaureate recipients from the 2015-16 academic year (B&B:16/17) will be entering full-scale collection in 2017 with a field test that was just completed (see 1850-0926). B&B cohorts prior to B&B:16 were approved under OMB# 1850-0729.

1. Respondent Universe

The respondent universe for the full-scale B&B study consists of all persons who completed requirements for the bachelor's degree during the 2007-08 academic year, and received their degree by June 30, 2009. These respondents will be surveyed for B&B in 2018. For the B&B field test, the respondent universe is the same except that the sampling year is 2006-07 and the survey year is 2017.

2. Statistical Methodology

a. Field Test Design

The B&B:08/18 field test will be implemented to fully test all procedures, methods, and systems planned for the B&B:08/18 full-scale study, in a realistic operational environment, prior to implementing them in the full-scale study. The field test is designed to test and validate data collection and monitoring procedures that will obtain the most accurate data in the least amount of time. Specific field test evaluation goals include the following:

- Identifying problematic data elements in the B&B:08/18 student interview;
- Determining optimal formats for questions for which multiple response options apply (check all options);
- Evaluating the willingness of sample members to submit resumes and allow LinkedIn account access in order to examine data quality, assess if it is feasible to reduce burden in the student interview, and to improve the overall efficiency of B&B data collections;
- Assessing varying approaches to introductory materials, source and signatory of emails, and the method by which data security is achieved; and
- Evaluating the representativeness of collected data when a paper survey mode is offered in addition to the online web and Computer-Assisted Telephone Interviewing (CATI) options.

Additionally, we will evaluate the time required to complete the interview, and sections of the interview, in order to identify possible instrument modifications that will save time in the full-scale interview. We will also conduct a reliability re-interview to evaluate the temporal consistency of selected interview items.

The initial field test sample for B&B:08/18 consisted of all interview respondents from the NPSAS:08 field test who completed requirements for their bachelor's degree at any

time between July 1, 2006, and June 30, 2007, and received their degree by June 30, 2008 (for the full-scale study, bachelor's degree requirements must have been completed between July 1, 2007 through June 30, 2008, and sample members must have received their degree by June 30, 2009). The NPSAS:08 field test yielded 2,460 interview respondents who indicated in the NPSAS field test interview that they had completed or expected to complete the requirements for their bachelor's degree in the 2006-07 academic year. The B&B eligible sample was then followed up in 2008 as the field test for B&B:08/09, after which 1,819 confirmed and potentially eligible sample members were retained for the B&B:08/12 field test in 2011. Following that interview, 1,588 sample members remained eligible. The sample for B&B:08/18 field test interview will include those 1,588 sample members, minus 8 sample members found to be either ineligible or deceased since 2011, and 20 who have never participated in a B&B:08 series interview and for whom insufficient data exist to determine their cohort eligibility. Consequently, the final B&B:08/18 field test sample size will be 1,560. Table 1 presents the field test sample counts by their survey response status through the B&B:08 interview series.

Table 1. Distribution of the B&B:08/18 field test and full-scale sample by interview response status for NPSAS:08, B&B:08/09, and B&B:08/12, among NPSAS:08 study members¹

NPSAS:08	B&B:08/09	B&B:08/12	Count
Field test collection			
Total			1,560
Respondent	Respondent	Respondent	801
Respondent	Respondent	Nonrespondent	132
Respondent	Nonrespondent	Respondent	106
Respondent	Nonrespondent	Nonrespondent	110
Nonrespondent	Respondent	Respondent	142
Nonrespondent	Respondent	Nonrespondent	67
Nonrespondent	Nonrespondent	Respondent	74
Nonrespondent	Nonrespondent	Nonrespondent	128
Full-scale collection			
Total			17,040
Respondent	Respondent	Respondent	13,330
Respondent	Respondent	Nonrespondent	1,476
Respondent	Nonrespondent	Respondent	996
Respondent	Nonrespondent	Nonrespondent	807
Nonrespondent	Respondent	Respondent	160
Nonrespondent	Respondent	Nonrespondent	53
Nonrespondent	Nonrespondent	Respondent	73
Nonrespondent	Nonrespondent	Nonrespondent	145

NOTE: All of the NPSAS:08 interview nonrespondents we are fielding were study members and, therefore, have some NPSAS data.

b. Full-scale Design

The sample design for the B&B:08/18 full-scale study is expected to mimic the plan for the field test described above. The sample will include all B&B:08/12 eligible sample members, but only those who were NPSAS:08 study members will be fielded.¹ NPSAS:08 non-study members will be counted among the nonrespondents in B&B:08/18. Based on historic response rates from this cohort, we are targeting a response rate of about 87

¹A sample member was classified as a *study member* if data were available for him or her on a set of key variables. Those variables, identified across the student interview, student records, and administrative data, were selected because they support the analytic objectives of the study.

percent among the eligible sample members, which will yield about 14,995 responding baccalaureate recipients. Sample size and response rates are subject to change based upon field test results.

3. Methods for Maximizing Response Rates

Response rates in the B&B:08/18 field test and full-scale data collections are a function of success in two basic activities: identifying and locating the sample members involved, then contacting them and gaining their cooperation. The following sections outline methods for maximizing response to the B&B:08/18 student survey.

1. Tracing of Sample Members

To achieve the desired response rate, we propose an integrated tracing approach designed to yield the maximum number of locates with the least expense. During the field test, we will evaluate the effectiveness of these procedures for the full-scale study effort. The steps of our tracing plan include the following elements.

- An advance tracing stage will include tracing steps taken prior to the start of data collection. These include batch database searches and advance intensive tracing (if necessary). Some cases will require more advance tracing before mailings can be sent or the cases can be worked in CATI. To handle cases for which mailing address, phone number, or other contact information is invalid or unavailable, B&B staff plan to conduct advance tracing of the cases prior to lead letter mailing and data collection. As lead information is found, additional searches will be conducted through interactive databases to expand on leads found.
- Data collection mailings and emails will be used to maintain persistent contact with sample members as needed, prior to and throughout data collection. Initial contact letters will be sent to parents and sample members prior to the start of data collection once OMB clearance is received. The letter will remind sample members of their inclusion in the study and request that they provide any contact information updates. Following the initial contact letters, an announcement letter will be sent to all sample members by the first day of data collection to announce the start of data collection. The data collection announcement will include a toll-free number, the study website address, a Study ID and password, and will request that sample members complete the web survey. Two days after the data collection announcement mailing, an email message mirroring the letter will also be sent to sample members.
- The telephone locating and interviewing stage includes calling all available telephone numbers and following up on leads provided by parents and other contacts.
- The pre-intensive batch tracing stage consists of the LexisNexis SSN and Premium Phone batch searches that will be conducted between the telephone locating and interviewing stage and the intensive tracing stage.
- The intensive tracing stage consists of tracers conducting database searches after all current telephone numbers have been exhausted. In B&B:08/09, about 77 percent of sample members requiring intensive tracing were located, and about 46 percent of those located responded to the interview. In B&B: 08/12, about 89 percent of sample members requiring intensive tracing were located, and about 39 percent of those located responded to the interview. Intensive interactive tracing differs from batch tracing in that a tracer can assess each case on an individual basis to determine which resources are most appropriate and the order in which they should be used. Intensive interactive tracing is also much more detailed due to the personal review of information. During interactive tracing, tracers utilize all previously obtained contact

information to make tracing decisions about each case. These intensive interactive searches are completed using a special program that works with RTI's case management system to provide organization and efficiency in the intensive tracing process. Sources that may be used, as appropriate, include credit database searches, such as Experian, various public websites, and other integrated database services.

- Other locating activities will take place as needed, including a LexisNexis email search conducted for nonrespondents toward the end of data collection.

2. Training for Data Collection Staff

Telephone data collection will be conducted at the contractor's call center. B&B staff at the call center will include Performance Team Leaders (PTLs) and Data Collection Interviewers (DCIs). Training programs for these staff members are critical to maximizing response rates and collecting accurate and reliable data.

Performance Team Leaders, who are responsible for all supervisory tasks, will attend project-specific training for PTLs, in addition to the interviewer training. They will receive an overview of the study, background and objectives, and the data collection instrument through a question-by-question review. PTLs will also receive training in the following areas: providing direct supervision during data collection; handling refusals; monitoring interviews and maintaining records of monitoring results; problem resolution; case review; specific project procedures and protocols; reviewing CATI reports; and monitoring data collection progress.

Training for DCIs is designed to help staff become familiar with and practice using CATI case management system and the survey instrument, as well as to learn project procedures and requirements. Particular attention will be paid to quality control initiatives, including refusal avoidance and methods to ensure that quality data are collected. DCIs will receive project-specific training on telephone interviewing and answering questions from web participants regarding the study or related to specific items within the interview. At the conclusion of training, all B&B call center staff must meet certification requirements by successfully completing a certification interview. This evaluation consists of a full-length interview with project staff observing and evaluating interviewers, as well as an oral evaluation of interviewers' knowledge of the study's Frequently Asked Questions.

3. Case Management System

Interviews will be conducted using a single web-based survey instrument for both web (including mobile devices) and CATI data collection. The data collection activities will be accomplished through a CATI case management system, which is equipped with the numerous capabilities, including: on-line access to locating information and histories of locating efforts for each case; a questionnaire administration module with full "front-end cleaning" capabilities (i.e., editing as information is obtained from respondents); sample management module for tracking case progress and status; and automated scheduling module which delivers cases to interviewers. The automated scheduling module incorporates the following features:

- *Automatic delivery of appointment and call-back cases at specified times.* This reduces the need for tracking appointments and helps ensure the interviewer is punctual. The scheduler automatically calculates the delivery time of the case in reference to the appropriate time zone.
- *Sorting of non-appointment cases according to parameters and priorities set by project staff.* For instance, priorities may be set to give first preference to cases

within certain sub-samples or geographic areas; cases may be sorted to establish priorities between cases of differing status. Furthermore, the historic pattern of calling outcomes may be used to set priorities (e.g., cases with more than a certain number of unsuccessful attempts during a given time of day may be passed over until the next time period). These parameters ensure that cases are delivered to interviewers in a consistent manner according to specified project priorities.

- *Restriction on allowable interviewers.* Groups of cases (or individual cases) may be designated for delivery to specific interviewers or groups of interviewers. This feature is most commonly used in filtering refusal cases, locating problems, or foreign language cases to specific interviewers with specialized skills.
- *Complete records of calls and tracking of all previous outcomes.* The scheduler tracks all outcomes for each case, labeling each with type, date, and time. These are easily accessed by the interviewer upon entering the individual case, along with interviewer notes.
- *Flagging of problem cases for supervisor action or supervisor review.* For example, refusal cases may be routed to supervisors for decisions about whether and when a refusal letter should be mailed, or whether another interviewer should be assigned.
- *Complete reporting capabilities.* These include default reports on the aggregate status of cases and custom report generation capabilities.

The integration of these capabilities reduces the number of discrete stages required in data collection and data preparation activities and increases capabilities for immediate error reconciliation, which results in better data quality and reduced cost. Overall, the scheduler provides an efficient case assignment and delivery function by reducing supervisory and clerical time, improving execution on the part of interviewers and supervisors by automatically monitoring appointments and call-backs, and reducing variation in implementing survey priorities and objectives.

Completed paper surveys will be returned to survey staff using an addressed, postage paid envelope enclosed in the package sent to sample members. Because use of the 10-item paper survey is experimental and offered to a small percentage of the field test sample, we are expecting that only a small number of completed paper surveys will be returned. Consequently, survey data will be keyed by telephone interviewers already working on the B&B:08/18 field test. Paper surveys can be scanned by call center staff if the mode is offered during the full-scale data collection; a description of those processes will be provided in the full-scale package.

4. Survey Instrument Design

Student interview preparation involved a meeting of the B&B technical review panel in November 2016. The focus of this first meeting was to discuss the key data elements to be included in the field test interview data collection, where employment outcomes of bachelor's degree recipients and their consideration of a career in K-12 teaching and STEM occupations feature prominently. The second meeting of the panel will take place in late fall of 2017 to consider content for the full-scale survey and the data collection design.

The B&B:08/18 student interview employs a web-based instrument and deployment system which has been in use since NPSAS:08. The system provides multimode functionality that can be used for self-administration, including on mobile devices, CATI, CAPI, or data entry. The instrument is provided in appendix E.

In addition to the functional capabilities of the case management system and web instruments described above, our efforts to achieve the desired response rate will include using established procedures proven effective across our large-scale studies. These include:

- Providing multiple response modes, including mobile-friendly self-administered and interviewer-administered options.
- Offering incentives to encourage response (see incentive structure described in Section 4, Tests of Procedures and Methods).
- Assigning experienced CATI interviewers who have proven their ability to contact and obtain cooperation from a high proportion of sample members.
- Training the interviewers thoroughly on study objectives, study population characteristics, and approaches that will help gain cooperation from sample members.
- Maintaining a high level of monitoring and direct supervision so that interviewers who are experiencing low cooperation rates are identified quickly and corrective action is taken.
- Making every reasonable effort to obtain an interview at the initial contact, but allowing respondent flexibility in scheduling interview appointments.
- Thoroughly reviewing all refusal cases and making special conversion efforts whenever feasible (see next section).

5. Refusal Aversion and Conversion

Recognizing and avoiding refusals is important to maximize the response rate. We will emphasize this and other topics related to obtaining cooperation during interviewer training. PTLs will monitor interviewers intensely during the early days of outbound calling and provide retraining as necessary. In addition, the supervisors will review daily interviewer production reports produced by the CATI system to identify and retrain any data collectors who are producing unacceptable numbers of refusals or other problems.

Refusal conversion efforts will be delayed for at least one week to give the respondent time after the initial refusal. Attempts at refusal conversion will not be made with individuals who become verbally aggressive or who threaten to take legal or other action. Refusal conversion efforts will not be conducted to a degree that would constitute harassment. We will respect a sample member's right to decide not to participate and will not impinge this right by carrying conversion efforts beyond the bounds of propriety. Sample members who explicitly refuse participation in the survey will be offered the mini survey option, described below.

4. Tests of Procedures and Methods

The B&B:08/18 field test will include two data collection experiments: the first set of experiments focuses on survey participation and aims at reducing nonresponse error, while the second experiment focuses on minimizing measurement error to further improve data accuracy. We will also conduct an observational study of sample members' willingness to provide resumes.

a. Data Collection Experiments

Experiment #1a: Tailoring of Contact Materials

Tailoring communications to encourage survey participation has long been done by interviewers (Groves and McGonagle 2001). Persuasion theory argues that interviewers who can successfully draw connections between survey participation and an

individual's attitude and beliefs are more successful in gaining compliance due to an individual's need for consistency (Groves et al. 1992). Furthermore, social exchange theory posits that individuals will comply with a request from strangers if they trust that the expected rewards will exceed the anticipated costs of complying (Dillman et al. 2014, 2017). It argues for tailoring and developing survey procedures that encourage positive social exchange and participation by highlighting the benefits and relevance associated with participation.

Building on these ideas, we propose to customize contact materials sent to study members to make the survey more attractive to different sample cases based on information available on the frame. Tailoring the advance letter provides a direct effort at persuasion to increase response rates and representativeness by directly targeting the decision-making strategy of respondents and increasing perceived relevance, topic saliency, and commitment (Groves et al. 1992). The limited empirical evidence supports these arguments. Compared to standard letters, for example, targeted letters improve response rates significantly among groups with lower-response propensity in an innovation panel (Lynn 2016).

Sample members will be assigned at random into one of two conditions described below.

1. *Treatment Group*: This targeted solicitation effort would personalize our participation requests, thereby conveying the value of their input to potential respondents. For example,
 1. "B&B is interested in your personal experience since you graduated from college during the 2007-08 school year.
 2. Data collected from B&B will help researchers and policymakers better understand how earning a bachelor's degree **in <MAJOR>** impacts choices for additional education and employment paths."
2. *Control Group*: This group would receive the traditional B&B contact and reminder materials using a generic, inviting language to emphasize the benefit of the survey participation. For example,
 3. "B&B is interested in your personal experience since you graduated from college during the 2007-08 school year.
 4. Data collected from B&B will help researchers and policymakers better understand how earning a bachelor's degree impacts choices for additional education and employment paths."

We will evaluate both groups based on response rates and measures of representativeness, for example, using R-indicators or Mahalanobis distance measures. R-indicators are random variables that measure the extent to which respondents are representative of the sample as a function of a set of available auxiliary variables. They are based on the sample standard deviation and the sample variance of the estimated response probabilities and provide insight into the possible risk of nonresponse bias of the response mean of survey items. The Mahalanobis calculation is based on the multivariate distance between a baseline respondent average and an individual nonrespondent value, calculated at multiple time points during data collection. Nonrespondents with high Mahalanobis distances can be considered most likely to contribute to nonresponse bias, and a reduction in the average value among nonrespondents may indicate that the data collection protocol is helping to obtain responses from cases with higher Mahalanobis values.

If the proposed intervention is successful in encouraging response to the field test survey, we will implement this strategy in the B&B:08/18 full scale data collection and explore its merit for other NCES surveys. We will carefully monitor response rates and representativeness in both groups. If there is a clearly superior strategy, we may end the experiment prematurely in favor of the more effective approach.

Experiment #1b: Emphasis of NCES as Source and Signatory of Emails

Generally, individuals “are more likely to comply with a request if it comes from an authority” (Groves et al. 1992, p. 472). This is based on an increased sense of legitimacy for certain research (i.e., the government needs this information) and on trust, due to government employees facing high penalties when disclosing provided information (Dillman et al. 2014). Furthermore, leverage-salience theory suggests that positive “attitudinal states [towards the sponsor] can increase the likelihood to participate if survey sponsor is positively valued and sponsorship is made salient during the survey request” (Groves et al. 2012). Positive institutional recognition on the outside of the envelope or in the identity of the sender of an email might increase the likelihood of sample members opening the letter/email (Dillman et al. 2014). Positive effects on response rates have been reported for legitimate organizations, such as a university, compared to other, unknown organizations (Groves et al. 2012; Avdeyeva and Matland 2013; Edwards et al. 2014).

While B&B is a survey conducted by the National Center for Education Statistics (NCES) in the U.S. Department of Education (ED), contractors have typically used their own or a study-specific email and telephone number to contact and support sample members. Drawing on insights from the literature, we propose to use federal email addresses that emphasize the NCES/ED role in conducting the survey. More specifically, sample members will be assigned at random into one of two conditions described below.

1. *Treatment Group*: This group of respondents will receive emails from an [###@ed.gov](#) email address, signed by the NCES study director. Otherwise, the content of the email will be the same as that sent to all other sample members.
2. *Control Group*: This group of respondents will receive emails from [BBemail@rti.org](#) email address, signed by the RTI study director. Otherwise, the content of the email will be the same as that sent to all other sample members.

All hard copy mailings will continue using envelopes and postcards with an ED return address.

We expect that emphasizing that the study is conducted by NCES/ED and sending emails from an [###@ed.gov](#) email address will result in a higher response rate and a more representative sample reducing the potential for nonresponse bias. We will evaluate both groups based on response rates and measures of representativeness such as R-Indicators or Mahalanobis distance measures. If the proposed intervention is successful in encouraging response to the field test survey, we will implement this strategy in the B&B:08/18 full scale data collection and explore its merit for other NCES surveys.

Experiment #1c: Security Questions

The promise of confidentiality is crucial to avert refusals (AAPOR 2014). In light of recent security breaches and given that information from prior survey participation is preloaded in the B&B:08/18 instrument, NCES intends to increase security by adding a series of questions to the survey that only the respondent would know. Adding security questions, however, can add burden to the instrument and may block legitimate respondents from gaining access to the survey, thereby decreasing participation. This

experiment is designed to test the impact of adding the questions to the survey. More specifically, we will compare two different formats of these security questions to which participants will be randomly assigned.

The first format will ask respondents to correctly identify the institution they attended to obtain their bachelor's degree from a list of 5 options. Only if the selected response is correct will the sample member continue into the interview. The second format will use the fact that we have social security numbers for the majority of the sample members. Sample members in this condition will receive the question to confirm the last four digits of their SSN. If the digits match, the interview continues. If the digits do not match, the sample member will be asked to correctly identify their previous address from a selection of addresses. Again, if this is done correctly they proceed to the interview. If the address selection results in an error, the sample member will be asked to correctly identify their telephone number from a selection of telephone numbers. If this last step fails as well, the interview ends and the sample member will be instructed to contact the helpdesk.

Each format has different advantages and disadvantages. The advantage of the first format is that the response burden is relatively low for the sample member. Yet, the degree of protection is not as high. While the second format provides the highest degree of protection, asking for the last four digits of the SSN is a potentially more sensitive question that might lead to breakoffs and induce nonresponse bias. Dillman et al. (1993), for example, show that response rates were significantly lower when asking for the entire SSN (3.4 percentage points from 71.4% to 68.0%) in a census mail survey that asked for the SSN for each member of the household (which might confound 'not willing' with 'not knowing' and overestimate the negative effects). To desensitize this request, we can remind respondents that we have their SSN on file and only use it to verify and protect their identity.

We will evaluate each strategy in terms of response rates, representativeness, the number of (non)matches, breakoffs, and timing. The results of this experiment will inform the B&B:08/18 full scale design.

Experiment #1d: Mini Surveys

Retreating to a shorter interview after a refusal has been successfully used to increase compliance rates in interviewer administered surveys compared to providing a shorter questionnaire from the start (Mowen and Cialdini 1980). In addition to offering shorter interviews, offering multiple modes of response may improve response rates and representativeness (e.g., Shettle and Mooney 2009; Dillman et al. 2014). Switching from web to mail, for example, has been shown to increase response rates between 4.7 to 19 percentage points and to increase representativeness by bringing in different types of respondents (e.g., Messer and Dillman 2011; Millar and Dillman 2011).² In a non-experimental design, Biemer et al. (2016) show that a mode switch from web to mail improved response rates by up to 20 percentage points.

The shorter the stated length of a survey, the lower the perceived burden for the respondent and the higher the expected benefits for the respondents. Motivated by this approach, we plan to offer an extremely abbreviated questionnaire – a mini survey – to all nonrespondents late in the data collection period. Obtaining information on these

² It is worth noting that these studies are lacking a true control condition, i.e., no web condition was followed-up in web only. The mode switch was from web to mail, or vice versa, hence number of contact attempts and mode switch are potentially confounded. We will extrapolate what the response rates would have been, had we not offered the mini survey, and compare the pure mini survey effect to that simulation. This allows us to test the mini survey effect and the mode switch effect.

nonrespondents is crucial to assess and increase sample representativeness (e.g., Kreuter et al. 2010). This request will be accompanied with a \$20 incentive, paid conditional upon participation by check or PayPal, and a request to upload their resumes for an additional \$10 incentive, conditional upon upload, also paid by check or PayPal. We expect that more sample members will complete the mini survey compared to the traditional abbreviated interview as the mini survey is less burdensome as it is even shorter. We expect it to yield higher participation rates and a more representative sample (e.g., Groves et al. 1992; Galesic and Bosnjak 2009). However, we expect lower resume upload rates compared to respondents completing the full survey, as these are hard to get respondents who will likely be less motivated to comply with the additional request.

Catering to respondents with different mode preferences is likely to increase survey participation and representativeness. Motivated by this approach, we plan to conduct a mode switch experiment, allowing nonrespondents to also complete the mini survey on paper (PAPI). Mailing out a questionnaire along with the mini survey request will be beneficial in multiple ways: 1) it will increase the likelihood of sample members opening the letter due to the different format and weight of the letter; 2) it will give sample members a chance to see the actual survey without having to go online, thereby giving them a realistic assessment of the actual, reduced response burden and showing them how easy it is to participate, instead of relying on a promise; 3) including a professional looking survey potentially increases the legitimacy of the request, and 4) allowing for mail as an additional mode is expected to change the likelihood of responding and bring in different respondents (e.g., respondents suspicious of clicking on links). We plan to randomly assign half of the late survey nonrespondents to the mini survey (plus resume request) and staying with the original survey modes (henceforth referred to as mini-web). The remaining half will receive the same request but will be given the opportunity to complete the mini survey using paper and pencil in a booklet that is mailed out along with the mini survey request (including a postage paid return envelope; referred to as mini-PAPI). Respondents in that condition will still have the option to complete the survey online, if preferred, and to upload a resume online. Respondents who choose to return the questionnaire by mail will be sent a thank you letter mentioning the opportunity to also share their resume.

The mini survey itself will consist of a maximum of 10 questions selected to optimize imputation and nonresponse bias assessments (listed in table 1).

Table 2: Items selected for the mini survey

Topic	Item
Post-BA Education	B18CPSTGRD
Employment	B18DNUMEMP; B18DEMPFTPTM; B18DEND01; B18DUNCM
Teaching	B18EANYTCH
Background	B18FHSCDR; B18AMARR; B18FDEPS; B18FDEP2

These efforts should lead to an increase in response rates and a potential reduction in nonresponse bias. More specifically, all late survey nonrespondents will be randomly assigned to one of two mini survey conditions:

1. *Treatment Group 1a.* Mini survey plus resume request in original modes (mini-web).
2. *Treatment Group 1b.* Mini survey plus resume request with PAPI mode (mini-PAPI)

In the final week or so of data collection all remaining nonrespondents will receive a final request to upload their resume as a nonresponse conversion attempt. This attempt will be incentivized with \$20. This will be the final contact for sample members.

We will investigate the increase in participation rates and representativeness when including mini survey respondents overall and by treatment groups (mini plus resume in either the original mode or the paper option) by tracking daily submission rates. We will also investigate respondents' willingness to comply with the resume upload throughout data collection. If the proposed intervention is successful in encouraging response to the field test survey, we will implement this strategy in the B&B:08/18 full scale data collection and explore its merit for other NCES surveys.

b. Experiment #2: Questionnaire Design

The B&B:08/18 field test questionnaire contains a set of questions for which respondents are asked to select items that apply to them from a list of options. B&B:08/18 currently employs two formats to ask these questions: 1) the traditional check-all format in which respondents are asked to check a box for each item that applies to them, and 2) the so-called forced-choice format that presents respondents with explicit yes-no options for each item in the list.

The common format in web surveys for asking these questions is check-all. However, check-all formats are awkward to administer over the telephone and are often considered to provide inferior data quality as they might encourage weak satisficing (e.g., Smyth et al. 2006). Furthermore, an unchecked box in a check-all format is hard to interpret as it might mean that 1) the response option does not apply, 2) the respondent missed the item in the list, or 3) that the respondent was unsure. Experimental studies suggest that forced-choice formats yield consistently higher endorsement rates suggesting deeper cognitive processing and higher data quality (Callegaro et al. 2015). This effect tends to be more pronounced in telephone surveys.

A recent meta-analysis investigating studies that compare both formats, questions the validity of these conclusions as potentially two competing mechanisms--with vastly different implications for data quality--are consistent with higher endorsement rates (Callegaro et al. 2015). The first mechanism is that the forced-choice format fosters deeper cognitive processing of each item and is thus associated with a higher data quality. The second mechanism is based on acquiescence bias that is associated with higher endorsement rates in the forced-choice format but implies lower data quality.

Validation studies on this topic are few and yield inconclusive evidence as to which format provides the 'best' data quality (see Callegaro et al. 2015). Given the use of both, check-all and forced-choice question formats in the B&B:08/18, further investigations are warranted. We suggest the following experimental design to investigate which of the above mechanisms is leading to these results and which question format yields the highest quality data. This experiment will inform the B&B:08/18 full scale but will also serve as a reference for other NCES studies.

Sample members will be assigned at random (upon login) to either the check-all design or the forced-choice design. We include one question with items that can be validated with auxiliary information to assess response accuracy across question formats. This allows us to disentangle acquiescence and deeper cognitive processing directly. For items that cannot be validated, we need an additional experiment to investigate which mechanism is at work. Respondents in the forced-choice condition will be randomized into two groups. The first forced-choice group sees 'yes-no' and the second sees the

reversed response options ‘no-yes.’ In order to avoid order effects in both formats and across survey modes the item order within each question will be randomized.

Respondents will be assigned at random into one of the following three conditions (upon login) applied to those questions that the majority of respondents are expected to see (for reasons of statistical power) and which have more than five items to choose from:

Treatment Group 1a: Sample members will receive the forced-choice yes-no format.

Treatment Group 1b: Sample members will receive the forced-choice no-yes format.

Control Group: Sample members will receive the check-all format.

The items selected for the experiment are provided in table 3. Please note that there is no item overlap between these items and those selected for the mini survey in experiment 2.

Table 3: Items selected for the forced choice experiment

Topic	Item
Post-BA Education	B18CFINAIDG01
Employment	B18DCHNG01
Background	B18FMILIT*; B18AHCOMP; B18FAFFCOST; B18FRETIR

*Note: NPSAS:16 is currently investigating the potential to validate B18FMILIT. If successful, this will be the validation item.

The amount of information collected will be the same for all groups. The three groups will be evaluated on endorsement rates, item nonresponse rates, and timing. The results of this experiment will inform the B&B:08/18 full scale instrument design.

c. Observational Study of Sample Members’ Willingness to Provide a Current Resume

Resume collection efforts have not been included in past B&B studies. Resumes can provide a useful data source to: 1) explore ways to reduce burden for future students; 2) provide other data as a potential alternative or complement for survey data (explore potential for imputation and to assess survey data quality); and 3) as a means to assess nonresponse bias and adjust for nonresponse (see below). To explore the value and utility of resumes for B&B:08/18 in general, we plan to ask respondents to upload their resume at the end of the survey. The rationale for asking for resumes at the end of the survey instead of the beginning is that: 1) we have built rapport (“foot-in-the-door approach,” see Freedman and Fraser 1966) and 2) that we are not sure how this request will be perceived and do not want to risk survey breakoffs. Furthermore, respondents might perceive the additional request to ask for resumes as tipping the burden-to-benefit ratio.

We propose to offer an additional \$10 incentive to thank survey respondents for the additional time and effort required by the resume upload process. A similar approach is taken by studies asking for additional information, such as biomarkers. Several studies show that incentives in web surveys in general appear to increase the likelihood of response with no negative impact on data quality (e.g., meta-analysis by Singer and Ye 2013; Medway and Tourangeau 2015). In addition, we will use resume requests as a means of nonresponse conversion by offering survey nonrespondents a \$20 incentive to upload a current resume by the end of field test data collection. We hope to get sample members/nonrespondents to provide their resume even if they are not especially

interested in the topic of the survey. This measure could reduce the potential for nonresponse bias (Groves et al. 2000; Groves 2006).

As part of field test data analysis, we will evaluate the quality of the data collected via resumes. To do this, we will compare resume data to information obtained in the survey and administrative records. Additionally, we can contrast cost of collecting resumes per case with the additional gains. The results of the B&B:08/18 field test collection will help inform the full scale data collection and other NCES surveys. We will report our findings to OMB in the B&B:08/18 full-scale submission.

d. Experimental Design

The experiments described above will test the numerous hypotheses outlined below. The experimental design includes estimation of the minimum differences between the control and treatment groups necessary to detect statistically significant differences. All experiments will be fully crossed.

The control and treatment groups with the null hypotheses to be tested are defined as follows:

Experiment #1a: Tailoring of Contact Materials

Sample members will be randomly assigned to one of two conditions (30:70 ratio tailored vs. generic).

- *Treatment Group:* Sample members will receive the tailored contact materials
- *Control Group:* Sample members will receive the generic contact materials.
 - 1a. Response rates will not be higher for the treatment group compared to the control group.
 - 1b. Representativeness: Respondents in the treatment group will not differ from respondents in the control group on key distributions and representativeness indicator values.

Experiment #1b: Emphasis of NCES as Source and Signatory of Electronic Mail

Sample members will be randomly assigned to one of two conditions (50:50 ratio).

- *Treatment Group:* Sample members will receive NCES emails.
- *Control Group:* Sample members will receive RTI emails.
 - 2a. Response rates will not be higher for the treatment group compared to the control group.
 - 2b. Representativeness: Respondents in the treatment group will not differ from respondents in the control group on key distributions and representativeness indicator values.

Experiment #1c: Security Questions

Sample members will be randomly assigned to one of two conditions (50:50 ratio).

- *Treatment Group 1a :* Respondents will receive the 'institution' security question.
- *Treatment Group 1b:* Respondents will receive the 'SSN-address-telephone number' security question.
 - 3a. Response rates will not be different in the 'institution' group compared to the 'SSN' group.
 - 3b. Representativeness: Respondents in the 'institution' group will not differ from respondents in the 'SSN' on key distributions and representativeness indicator values.

4. The number of (non)matching cases will not be different in the 'institution' group compared to the 'SSN' group.
5. The proportion of breakoffs will not be different in the 'institution' group compared to the 'SSN' group.
6. Mean time to complete these questions will not be different in the 'institution' group compared to the 'SSN' group.

Experiment #1d: Mini Surveys

Nonrespondents in the first 8 weeks of data collection will be randomly assigned to one of two conditions (50:50 ratio).

- *Treatment Group 1a*: Nonrespondents will receive the 'mini survey' (\$20 incentive) request followed by the request to upload their resume (\$10 incentive) in the original survey modes.
 - *Treatment Group 1b*: Nonrespondents will receive the 'mini survey' (\$20 incentive) request followed by the request to upload their resume (\$10 incentive) in the original survey modes and PAPI.
- 7a. Response rates will not be higher when including respondents who completed the mini survey.
 - 7b. Mini survey respondents and respondents completing the full survey do not differ on key distributions and representativeness indicator values.
 8. Mini survey respondents (incl. PAPI) will not have higher participation rates compared to mini survey respondents in the original survey modes.
 9. Asking mini survey respondents to upload their resumes will not lead to an increase in resume upload rates compared to the full survey and resume request.

Experiment #2: Questionnaire Design

Respondents will be randomly assigned to one of three conditions.

- *Treatment Group 1a*: Respondents will receive the forced-choice yes-no.
 - *Treatment Group 1b*: Respondents will receive the forced-choice no-yes.
 - *Control Group*: Respondents will receive the check-all format.
10. Mean time to complete these questions will not be higher for the treatment groups compared to the control group.
 11. Item missingness rate will not be different for the treatment groups than the control group.
 12. Endorsement rates for the positive response options will not be higher for the treatment groups than the control group.
 13. Endorsement rates for the positive response options will not be higher for the treatment group 1a than the treatment group 1b.

e. Detectable Differences

The differences between the control and treatment group(s) necessary to detect statistically significant differences are shown in Tables 4 and 5. The following assumptions were made in computing detectable differences.

- The statistical tests will have 80 percent power with an alpha of 0.05.
- Detectable differences with 95 percent confidence were calculated as follows:
 - o Hypotheses 1b, 2b, 3a - 5, 7b, 8, and 11 assume a two-tailed test;

- o The remainder assumes a one-tailed test.
- The sample will be unequally distributed across the experimental groups for hypotheses 1 and 2 (30:70 ratio).
- The sample will be equally distributed across the two/three experimental groups for the remaining hypotheses.
- The assumed conservative response rate is 60 percent out of 1,560 sample members.
- The assumed conservative response rate to the full survey is 50 percent; the response rate to the mini survey is expected to be 20 percent.
- The resume upload rate is expected to be 20 percent.
- The assumed average R value is 0.5 (Schouten et al. 2009) ranging between R=1 (if se=0), i.e., perfect representation, and R=0 (se=0.5).
- The match rate will be 0.9 in hypothesis 4 (corresponding to a 0.1 nonmatch rate).
- The breakoff rates will be 0.1 in hypothesis 5.
- The mean time to complete the security question will be 30 seconds.
- The expected proportion of resume uploads is 0.2.
- The mean time to complete each question will be 1 minute.
- The item missingness rate for the control group for hypothesis 11 will be 5 percent.
- The endorsement rate for each question is 0.5.

Table 4. Detectable differences (two group comparisons)

Hypothesis	Group 1		Group 2		Detectable difference, with 95 percent confidence
	Definition	Sample size	Definition	Sample size	
1a	Generic	468	Tailored	1,092	6.6%
1b	Generic	468	Tailored	1,092	7.7%
2a	RTI	780	NCES	780	6.8%
2b	RTI	780	NCES	780	7.1%
3a	Institutions	780	SSN	780	6.8%
3b	Institutions	780	SSN	780	7.1%
4	Institutions	433	SSN	433	5.0%
5	Institutions	433	SSN	433	6.4%
6	Institutions	433	SSN	433	11.4 sec.
10	Check-all	260	Forced-choice (1a, 1b)	520	11.3 sec.
11	Check-all	260	Forced-choice (1a, 1b)	520	5.9%
12	Check-all	260	Forced-choice (1a, 1b)	520	9.4%
13	Forced-choice (1a)	260	Forced-choice (1b)	260	10.8%

Table 5: Detectable differences (stacked method, dependent samples)

Hypothesis	Definition	Sample size	Detectable difference, with 95 percent confidence
7a	Mini survey	1,560	4.5%
7b	Survey-completes	1,560	5.0%
8	Mini survey (incl. PAPI)	1,560	3.7%
9	Mini survey resume upload	1,560	3.7%

5. Reviewing Statisticians and Individuals Responsible for Designing and Conducting the Study

The study is being conducted by NCES/ED. The following statisticians at NCES are responsible for the statistical aspects of the study: Mr. Ted Socha, Dr. Tracy Hunt-White, Dr. David Richards, Dr. Sean Simone, Dr. Elise Christopher, and Dr. Gail Mulligan. NCES's prime contractor for B&B:08/18 is the RTI International (RTI). The following staff members at RTI are working on the statistical aspects of the study design: Dr. Jennifer Wine, Ms. Melissa Cominole, Ms. Jennifer Cooney, Mr. Jeff Franklin, Dr. Antje Kirchner, Dr. T. Austin Lacy, Dr. Emilia Peytcheva, Mr. Peter Siegel, and Ms. Ashley Wilson.

Subcontractors include Coffey Consulting; Hermes; HR Directions; Research Support Services; Shugoll Research; and Strategic Communications, Inc. Consultants are Dr. Sandy Baum, Ms. Alisa Cunningham, and Dr. Stephen Porter. Principal professional RTI staff, not listed above, who are assigned to the study include Ms. Donna Anderson and Ms. Chris Rasmussen.

C. References

- Avdeyeva, O.A., & Matland, R.E. 2013. An Experimental Test of Mail Surveys as a Tool for Social Inquiry in Russia. *International Journal of Public Opinion Research*, 25(2), 173-194.
- Biemer, P., Murphy, J., Zimmer, S., Berry, C., Deng, G., Lewis, K. 2016. A Test of Web/PAPI Protocols and Incentives for the Residential Energy Consumption Survey. Paper presented at the 2016 Annual Conference of the American Association for Public Opinion Research, Austin, TX (May 13, 2016).
- Callegaro, M., Murakami, H., Tepman, Z., and V. Henderson. 2015. Yes-no Answers Versus Check-all in Self-administered Modes. *International Journal of Market Research*, 57(2): 203-223.
- Deming, W. E. (1953). On a Probability Mechanism to Attain an Economic Balance Between the Resultant Error of Response and the Bias of Nonresponse. *Journal of the American Statistical Association*, 48(264), 743-772.
- Dillman, D.A. , Sinclair, M.D., and Clark, J.R. 1993. Effects of Questionnaire Length, Respondent-Friendly Design, and a Difficult Question on Response Rates for Occupant-Addressed Census Mail Surveys. *Public Opinion Quarterly*, 57(3), 289-304.
- Dillman, D.A., Smyth, J.D., and Christian, L.M. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method 4th Edition*. John Wiley & Sons, Hoboken, NJ.
- Dutwin, D., Loft, J.D., Darling, J., Holbrook, A., Johnson, T., Langley, R.E., Lavrakas, P.J., Olson, K., Peytcheva, E., Stec, J., Triplett, T., and Zukerberg, A. 2014. Current Knowledge and Considerations Regarding Survey Refusals. AAPOR Task Force Report on Survey Refusal. Retrieved at http://www.aapor.org/Education-Resources/Reports/Current-Knowledge-and-Considerations-Regarding-Sur.aspx#_Toc393959577.
- Edwards, M.L., Dillman, D.A., and Smyth, J.D. 2014. An Experimental Test of the Effects of Survey Sponsorship on Internet and Mail Survey Response. *Public Opinion Quarterly*, 78(3), 734-750.
- Freedman, J.L. and Fraser, S.C. 1966. Compliance Without Pressure: The Foot-in-the-door Technique. *Journal of Personality and Social Psychology*, 4(2), 196-202.

- Galesic, M. and Bosnjak, M. 2009. Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. *Public Opinion Quarterly*, 73(2), 349-360.
- Groves, R.M. 2006. Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, 70(5), 646-675.
- Groves, R.M., Cialdini, R., and Couper, M. 1992. Understanding the Decision to Participate in a Survey. *Public Opinion Quarterly*, 56(4), 475-495.
- Groves, R.M., and McGonagle, K.A. 2001. A Theory-guided Interviewer Training Protocol Regarding Survey Participation. *Journal of Official Statistics*, 17, 249-65.
- Groves, R.M., Singer, E. and Corning, A. 2000. Leverage-Saliency Theory of Survey Participation. Description and Illustration. *Public Opinion Quarterly*, 64, 299-308.
- Groves, R.M., Presser, S., Tourangeau, R., West, B.T., Couper, M.P., Singer, E., and Toppe, C. 2012. Support for the Survey Sponsor and Nonresponse Bias. *Public Opinion Quarterly*, 76, 512-24.
- Kreuter, F., Olson, K., Wagner, J., Yan, T., Ezzati-Rice, T.M., Casas-Cordero, C., Lemay, M., Peytchev, A., Groves, R.M., and Raghunathan, T.E. 2010. Using Proxy Measures and Other Correlates of Survey Outcomes to Adjust For Non-Response: Examples from Multiple Surveys. *Journal of the Royal Statistical Society: Series A*, 173(2), 389-407.
- Medway, R.L. and Tourangeau, R. 2015. Response Quality in Telephone Surveys. Do Prepaid Incentives Make a Difference? *Public Opinion Quarterly*, 79(2), 524-543.
- Messer, B.L., and Dillman, D.A. 2011. Surveying the General Public Over the Internet Using Address-Based Sampling and Mail Contact Procedures. *Public Opinion Quarterly*, 75(3), 429-457.
- Millar, M.M., and Dillman, D.A. 2011. Improving Response to Web and Mixed-Mode Surveys. *Public Opinion Quarterly*, 75(2), 249-269.
- Mowen, J.C. and Cialdini, R.B. 1980. On Implementing the Door-in-the-Face Compliance Technique in a Business Context. *Journal of Marketing Research*, 17, 253-258.
- Lynn, P. 2016. Targeted Appeals for Participation in Letters to Panel Survey Members. *Public Opinion Quarterly*, 80(3), 771-782
- Schouten, B., Cobben, F. and Bethlehem, J. 2009. Indicators for the Representativeness of Survey Response. *Survey Methodology*, 35(1), 101-113.
- Shettle, C. and Mooney, G. 1999. Monetary Incentives in US Government Surveys. *Journal of Official Statistics*, 15(2), 231-250.
- Singer, E. and Ye, C. 2013. The Use and Effects of Incentives in Surveys. *Annals. Annals of the American Academy of Political and Social Science*, 645(1), 112-141.
- Smyth, J.D., Dillman, D.A., Christian, L.M., and Stern, M.J. 2006. Comparing Check-all and Forced-choice Question Formats in Web Surveys. *Public Opinion Quarterly*, 70(1), 66-77.