

2016/20 Baccalaureate and Beyond (B&B:16/20) Full-Scale Study

Supporting Statement Parts B and C

OMB # 1850-0926 v.11

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

December 2019
revised June 2020
revised August 2020

Contents

B. Collection of Information Employing Statistical Methods.....	1
1. Respondent Universe - B&B:16/20 Target Population.....	1
2. Statistical Methodology - B&B:16/20 Sample Design.....	1
3. Methods for Maximizing Response Rates.....	1
a. Tracing of Sample Members.....	2
b. Training for Data Collection Staff.....	2
c. Case Management System.....	3
d. Survey Instrument Design.....	4
e. Refusal Aversion and Conversion.....	4
4. Tests of Procedures and Methods.....	4
5. Reviewing Statisticians and Individuals Responsible for Designing and Conducting the Study.....	12
C. References.....	14

Tables

Table 1.....	B&B:16/17 full-scale study sample by eligibility status	1
Table 2.....	Sample composition by experimental condition	7
Table 3.....	B&B:16/20 full-scale data collection protocols by data collection phase and group assignment.....	8

B. Collection of Information Employing Statistical Methods

This request is to conduct the 2016/20 Baccalaureate and Beyond Longitudinal Study (B&B:16/20) full-scale data collection. A previous submission covered panel maintenance activities as well as administrative matching activities (1850-0926 v. 8). B&B:16/20 is the second follow-up of sample members from the 2015-16 National Postsecondary Student Aid Study (NPSAS:16) who were baccalaureate recipients during the 2015–16 academic year. For details on the NPSAS:16 sample and full-scale study design, see NPSAS:16 Full Scale (OMB# 1850-0666 v. 15-19) Supporting Statement Part B. B&B cohorts prior to B&B:16 are approved under OMB# 1850-0729 while the B&B:16 cohort is approved under OMB# 1850-0926.

1. Respondent Universe – B&B:16/20 Target Population

The target population for B&B:16/20 full-scale study includes all eligible NPSAS:16 sample members who completed requirements for the bachelor’s degree from NPSAS-eligible institutions during the 2015–16 academic year, that is, between July 1, 2015 and June 30, 2016, and were awarded their baccalaureate degree by the institution no later than June 30, 2017. There is a known and well-defined probability of selection for each student in the B&B sample. Through the institution awarding the degree, each completer has exactly one linkage to the B&B sampling frame.

The final weighted response rate for the most recent student survey of the B&B:16 cohort, during the first follow-up (B&B:16/17) full-scale data collection, was 69 percent. The final weighted response rate for the second follow-up of the previous, B&B:08 cohort, during the B&B:08/12 full-scale data collection, was 78.3 percent.

2. Statistical Methodology – B&B:16/20 Sample Design

The sample design for the B&B:16/20 full-scale study includes all eligible sample members from B&B:16/17. A sample member from B&B:16/17 was considered ineligible for the study if they had not completed their degree requirements for a bachelor’s degree at the institution they were sampled from in the NPSAS study during the 2015-15 academic year (July 1, 2015 through June 30, 2016), and was awarded the degree after June 30, 2017. Sample members identified as deceased during the B&B:16/17 data collection will also be excluded from the B&B:16/20 sample. All other previous B&B:16/17 sample members will be included in the B&B:16/20 sample regardless of prior response status. Table 1 shows the distribution of the B&B:16/17 sample by eligibility status to determine the B&B:16/20 sample.

Table 1. B&B:16/17 full-scale study sample by eligibility status

B&B:16/17 status	Count
Total	28,796
Eligible or unknown eligibility	26,509
Known ineligible*	2,259
Deceased	28

* Includes sample members who did not complete their degree requirements in the 2015-16 academic year, plus those who completed their degree requirements in the 2015-16 academic year at the NPSAS school but were not awarded a degree until after June 30, 2017.

In NPSAS:16, a sample member had to provide specific data elements through the student survey and administrative records in order to meet the minimum requirements to be a study member. Sample members who did not meet these requirements were considered non-study members. NPSAS:16 non-study members who were identified as potentially eligible for the B&B:16 cohort will be administered an eligibility screener in

B&B:16/20 but will be considered nonrespondents to the B&B:16/20 survey. This is similar to what was done in B&B:16/17. Since eligibility for the B&B:16 cohort has not been confirmed for all sample members, additional ineligible sample members may be identified in the course of the B&B:16/20 data collection.

3. Methods for Maximizing Response Rates

Achieving high response rates in the B&B:16/20 full-scale study data collection will depend on successfully identifying and locating sample members and being able to contact them and gain their cooperation. As was used successfully in B&B:08/12, shortly before the B&B:16/20 full-scale data collection begins, we will send an address update/initial contact mailing/email in the form of a greeting card to remind sample members, other than NPSAS:16 non-study members and double nonrespondents, of the study. The following sections outline additional methods for maximizing response to the B&B:16/20 data collection.

a. Tracing of Sample Members

To yield the maximum number of located cases with the least expense, we designed an integrated tracing approach with the following elements.

- Tracing activities conducted prior to the start of data collection will include batch database searches, such as to the National Change of Address (NCOA), for cases with enough contact information to be matched. To handle cases for which contact information is invalid or unavailable, B&B staff will conduct additional advance tracing through proprietary interactive databases to expand on leads found.
- Hard copy mailings, emails, and text messages will be used to maintain ongoing contact with sample members, prior to and throughout data collection. A panel maintenance mailing was sent in November 2019 to request that sample members update their contacting information (previously approved under the B&B:16/20 full-scale panel maintenance package, 1850-0926, v. 8). Panel maintenance materials are provided in appendix C, while the contacting materials for the full-scale study are provided in appendix E.
- At the start of data collection, greeting cards sent to sample members will request that they update their contact information. A follow-up reminder email will be sent approximately 2 weeks after the card to remind them to respond. Also, at the start of data collection, we will send a letter to announce the start of data collection. The announcement will include a request that sample members complete the web survey and will provide each sample member a Study ID and password, the study website address, and a toll-free number to the help desk. Sample members who did not complete either the NPSAS:16 or B&B:16/17 survey, and those who completed only an abbreviated survey for B&B:16/17, will receive \$2 cash (or paid by PayPal if a good address is not available) with the data collection announcement. After the data collection announcement mailing, an email message mirroring the letter will also be sent.
- The telephone locating and interviewing stage will include 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.
- Once all known telephone numbers are exhausted, a case will move into the intensive tracing stage during which tracers will conduct interactive database searches using all known contact information for a sample member. During the B&B:16/17 full-scale study, about 89 percent of sample members who reached the intensive tracing stage were located, and about 25 percent of those located responded to the survey. With interactive tracing, a tracer assesses each case on an individual basis to determine which resources are most appropriate and the order in which each should be used. 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.

b. 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.

PTLs, 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 reports generated from the ongoing Computer Assisted Telephone Interviewing (CATI); and monitoring data collection progress.

Training for DCIs is designed to help staff become familiar with and practice using the 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.

c. Case Management System

The B&B:16/20 survey will be conducted using a single web-based survey instrument for both web (including mobile devices) and CATI data collection. Data collection activities will be monitored through a CATI case management system, which is equipped with the numerous capabilities, including: online 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.

d. Survey Instrument Design

The survey will employ 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, Computer-Assisted Personal Interview (CAPI), or data entry.

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 in other large-scale studies we have completed. These include:

- Providing multiple response modes, including mobile-friendly self-administered and interviewer-administered options.
- Offering incentives to encourage response.
- 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 a completed interview at the initial contact while allowing respondent flexibility in scheduling appointments to be interviewed.
- Providing assurance of confidentiality procedures, including requiring respondents to answer security questions before obtaining and resuming access to the survey and the survey automatically logging out of a session after 20 minutes of inactivity.
- Thoroughly reviewing all refusal cases and making special conversion efforts whenever feasible (see next section).

e. Refusal Aversion and Conversion

Recognizing and avoiding refusals is important to maximize the response rate. We will cover 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.

4. Tests of Procedures and Methods

B&B:16/20 data collection, which will begin in summer 2020, will involve three distinct data collection groups and four main data collection phases. This general setup builds upon the designs implemented in B&B:16/17 and other B&B studies where it has contributed to maximizing response rates and minimizing the potential for nonresponse bias. In B&B:16/20 we plan to implement differential treatments based on prior round response status, an approach that was successfully implemented in the B&B:16/17 field test, where NPSAS:16 field test nonrespondents received either an aggressive or a default protocol. The response rate among NPSAS:16 field test nonrespondents who received the aggressive protocol was about 12% higher than the group that received the default protocol (37%; default 25% response rate $t(2,097) =$

3.52, $p < .001$). For the B&B:16/20 full-scale design, we will distinguish the following groups and design protocols:

NPSAS:16 non-study members and NPSAS:16 and B&B:16/17 survey nonrespondents: Sample members who were NPSAS:16 non-study members (i.e. sample members lacking enough information from the NPSAS survey and administrative collections to qualify them as NPSAS study members; $n=1,248$) will receive only an eligibility screener protocol. Sample members who failed to respond both to NPSAS:16 and B&B:16/17 ($n=2,609^1$), referred to as *double nonrespondents*, will also receive only an eligibility screener protocol. The goal of this treatment is to remove ineligible sample members from the response rate denominator.

NPSAS:16 or B&B:16/17 survey nonrespondents and B&B:16/17 abbreviated respondents: Sample members who either failed to respond to NPSAS:16 or B&B:16/17, referred to as *ever nonrespondents*, and respondents who completed the B&B:16/17 abbreviated survey will receive an aggressive data collection protocol ($n=7,622$). The goal of this treatment is to convert hard-to-get sample members as early in data collection as possible, based on prior evidence that the default data collection protocol is not successful for this group.

NPSAS:16 and B&B:16/17 survey respondent: Sample members who responded to both NPSAS:16 (full or abbreviated survey) and B&B:16/17 (full survey), referred to as *double respondents*, will receive a default data collection protocol ($n=15,030$).

In lieu of the B&B:16/20 field test, which is typically used to test data collection procedures and methodologies, we will conduct limited testing with a subset of the full-scale sample, a *calibration sample*. The calibration sample will be drawn from the B&B:16/20 full-scale sample in advance of the start of main sample data collection. Within the calibration sample, those sample members who were *ever nonrespondents* and B&B:16/17 abbreviated survey respondents will be sent a \$2 prepaid incentive with the announcement of the survey launch. The experiment to be conducted will seek to identify the better of two envelope designs for communicating presence of the cash in the envelope to sample members and, ultimately, increasing overall participation. Since data collected from the calibration sample will be included in the final data files, the calibration cases will otherwise follow the same data collection protocols as the main sample cases. The experiment is unlikely to have a direct effect on individual survey responses since it designed solely to impact sample members' likelihood to participate in the study. This is intentional; if the experiment produced large differences in the survey estimates, the data could not be combined with the main sample.

The following section describes the test of data collection procedures using the calibration sample, followed by a description of the main data collection procedures.

B&B:16/20 Calibration Experiment (Revised May 2020; Updated August 2020)

The goal of the B&B:16/20 Calibration Experiment was to test two alternative approaches to communicating the presence of the \$2 prepaid cash incentive to

¹ This is the overall number of double nonrespondents. However, double nonresponding cases that already completed the eligibility screener in B&B:16/17 will not be fielded again.

sample members in the aggressive data collection protocol²: (1) An envelope with a \$2 *Gift Enclosed. See details inside.* message and (2) an envelope that directly displays a proportion of a \$2 bill in the window of the envelope.

There is increasing evidence of mail theft as a result of the coronavirus pandemic stimulus checks by the Internal Revenue Service. Because of that, we think that showing cash in an envelope window or suggesting that cash is included is too risky, and we recommend altering the B&B:16/20 Calibration Experiment to focus on a test of two different prepaid incentive forms.

Cash prepaid incentives have been shown to significantly increase response rates in both interviewer-administered and self-administered surveys, reducing the potential for nonresponse bias (e.g., Church 1993; Cantor et al. 2008; Goeritz 2006; Medway and Tourangeau 2015; Messer and Dillman 2011; Parsons and Manierre 2014; Singer 2002). Evidence from the B&B:16/17 FT study, however, indicates that prepaid incentives sent via PayPal do not significantly increase response rates likely due to low acceptance rates. We hypothesize that this is primarily a result of sample members remaining unaware of the prepaid incentive payment: Sample members may fail to read the corresponding announcement in the contacting materials, they may only infrequently check their e-mail, or PayPal balances, and/or if they do check their balances, they may miss the prepaid incentive because of its small value.

The Calibration Experiment will investigate if an alternative form of communicating the prepaid PayPal incentive is associated with an increase in response rate, similar to the effect associated with cash prepaid incentives. Specifically, we suggest adding a separate index card announcing the \$2 prepaid PayPal incentive in the data collection announcement mailing while keeping all else equal (including later email communications). We hypothesize that the index card will stand out and make the \$2 prepaid payment more tangible, similar to how a \$2 bill would stand out. While the cost of the mailing materials is similar, sending out PayPal incentives is generally considered to be safer as it provides us with a way to track payments.

We recommend testing two approaches of communicating the presence of the \$2 prepaid incentive to the sample members in the aggressive protocol:

- Control Group will receive a mailing with a \$2 prepaid cash incentive enclosed.
- Treatment Group will receive a mailing with a \$2 prepaid PayPal incentive announced on a separate index card.

We will randomly select a calibration sample of 3,130 ever nonresponding sample members (i.e., either did not participate in the NPSAS:16 base year or the B&B:16/17 survey) or respondents who only completed the B&B:16/17 abbreviated survey, to receive either form of the prepaid incentive³. This sample will allow for comparisons of response rates among two equally-sized treatment groups of 1,565 sample members each and provide enough power to detect at least a 5 percentage point difference in response rates assuming 80 percent power, type I error of 5 percent, and a base response rate of 50 percent. This calculation assumes a 2-sided chi-square test of the response proportions.

The experiment described above will allow us to test the following hypotheses:

² To receive the prepaid incentive, sample members needed to have responded to either NPSAS:16 base year or the B&B:16/17 full survey or completed the B&B:16/17 abbreviated survey.

³ 96% of these cases have at least one good email address which can be used to send the PayPal prepaid incentive.

- H1. There is no statistically significant difference in response rates between Control Group and Treatment Group.
- H2. There are no statistically significant differences in representativeness (demographic characteristics) between Control Group and Treatment Group.

The proposed experimental period for the experiment is two weeks starting in early July 2020, after which we will analyze the results to determine which approach to recommend for the main data collection starting August 2020.

The results of the calibration experiment at the end of the experimental evaluation period are as follows.

Response Rates. Comparing the B&B:16/20 calibration sample response rates for the Control Group (cash; AAPOR RR1⁴=22.1 percent) and the Treatment Group (PayPal; 20.3 percent) using a two-tailed z-test yields no statistically significant differences in response rates between the two group ($z = -1.26, p = 0.21$). This finding is promising in that announcing the \$2 prepaid PayPal incentive using an index card that stands out produces similar response rates as a \$2 cash prepaid incentive.

Representativeness. In addition to monitoring response rates, we conducted nonresponse bias analyses to assess the representativeness of the responding sample for the cash and the PayPal group. Table 2 displays summary measures for the demographic distributions by group for the responding sample, as well as the overall sample including nonresponding cases. Comparing the responding sample composition with the overall sample composition shows the magnitude of nonresponse bias. For example, the overall sample in the Control Group consists of 57.4 percent females. At the end of the calibration evaluation period the responding sample overrepresents females by 7.7 percentage points with a total of 65.1 percent females.

The table shows that the two groups do not yield samples with a different demographic composition compared to their overall sample estimates and suggests no differential nonresponse bias except for age. A formal two-sided z-test shows that we fail to reject the null hypothesis of no difference in all instances so far except for age ($z = -2.38, p = 0.017$). The PayPal incentive is significantly more effective in the age groups around 29 and younger compared to 30 and older resulting in a significantly younger respondent sample in the Treatment Group.

⁴ Unless noted otherwise all response rates reported refer to the response rate 1 (RR1) as defined by the standards of the American Association for Public Opinion Research (AAPOR 2016). The RR1 is the number of complete interviews (excluding partial interviews) divided by the number of complete and partial interviews plus all non-interviews (excluding confirmed ineligible).

Table 2: Sample composition by experimental condition		
	Control Group	Treatment Group
	Cash	PayPal
Age (mean)		
Respondent Sample	31.4	29.8
Overall Sample (n=3,080) ¹	31.6	31.3
Female (in percent)		
Respondent Sample	65.1	61.6
Overall Sample (n=3,060) ¹	57.4	55.7
White (in percent)		
Respondent Sample	78.1	78.3
Overall Sample (n=3,110) ¹	73.2	74.0
Hispanic (in percent)		
Respondent Sample	14.1	12.4
Overall Sample (n=3,060) ¹	15.3	13.2
Employment (in percent)		
Respondent Sample	92.3	93.7
Overall Sample (n=1,920)	89.6	92.4

¹ Sample sizes for the overall differ due to missing data.
Note: Results exclude ineligible cases. Partial interviews are considered nonrespondents for analytic purposes.
Source: U.S. Department of Education, National Center for Education Statistics, 2016/20 Baccalaureate and Beyond (B&B:16/20)

Overall, while there is no statistically significant difference in response rates between the \$2 cash prepaid incentive and the \$2 PayPal prepaid incentive, there is a statistically significant difference in the resulting sample composition when it comes to age: The Treatment Group results in a statistically significantly younger respondent sample. Given the differential effectiveness of the PayPal incentive among the younger and older sample members, for the B&B:16/20 main data collection (Aggressive Protocol) we recommend proceeding with the incentive design for the Control Group (\$2 cash prepaid incentive) for individuals aged 30 and older and with the incentive design for the Treatment Group (\$2 PayPal prepaid incentive) for individuals aged 29 and younger.

B&B:16/20 Main Data Collection

As discussed above, the B&B:16/20 data collection will involve three distinct data collection groups and four data collection phases. Table 3 presents the type and timing of interventions to be applied in the main data collection by groups and protocol.

Except for the calibration experiment, phase duration will be decided based on phase capacity—the time at which a subgroup’s estimates remain stable regardless of additional data collection efforts. For example, during the Early Completion Phase, key metrics are continually monitored and when they stabilize over a period of time, cases are then transferred to the next phase. Phase capacity will be determined based on a series of individual indicators within each data collection protocol. For example, we will assess response rates and other level of effort indicators over time accounting for covariates such as control of institution, i.e., public, private nonprofit, or private for-profit.

Turning to incentives, the baseline incentive for the eligibility screener protocol will be \$5. The baseline incentive for the aggressive protocol will be \$35 in addition to a \$2 prepaid incentive, and possibly a \$5 early bird incentive or a \$10 incentive boost, discussed below. The maximum possible total incentive is \$47 in this aggressive data collection protocol. The baseline incentive for the default protocol will be \$30 with

either a \$2 prepaid incentive or a \$10 promised flash incentive as a nonresponse conversion strategy leading to a maximum possible total incentive of \$32 or \$40. Beyond the baseline incentives, both survey data collection protocols employ similar interventions, although the timing of these interventions differs across groups: interventions occur sooner in the aggressive protocol.

Table 3. B&B:16/20 full-scale data collection protocols by data collection phase and group assignment

	Data Collection (DC) Group Assignments		
	Group 1 (Eligibility Screener)	Group 2 (Aggressive Protocol)	Group 3 (Default Protocol)
Sample	NPSAS:16 Non-study Members (n=1,248) and Double Nonrespondents (n=2,609)	Ever Nonrespondents and B&B:16/17 Abbreviated Respondents (n=7,622)	Double Respondents (n=15,030)
Data Collection Protocols			
Prior to data collection (upon package approval)	N/A	<ul style="list-style-type: none"> Greeting card 	<ul style="list-style-type: none"> Greeting card
Early Completion Phase (July 2020)	<ul style="list-style-type: none"> Screener mail, email, and text invitation Mail, email, and text message reminders 	<ul style="list-style-type: none"> DC announcement mail, text, and email offering \$2 prepaid incentive and additional \$5 "early bird" incentive CATI starts 2 weeks after mail outs – continued through all phases 	<ul style="list-style-type: none"> DC announcement mail, text, and email Mode tailoring beginning with B&B:16/17 completion mode
Production Phase I (July 2020)	<ul style="list-style-type: none"> Mail, email, and text message reminders 	<ul style="list-style-type: none"> Postcard, email, and text message reminders – continued through all phases 	<ul style="list-style-type: none"> Light CATI outbound begins – continued through all phases Postcard, email, and text message reminders – continued through all phases
Production Phase II (January 2021)		<ul style="list-style-type: none"> Abbreviated survey offered 	<ul style="list-style-type: none"> As needed, \$2 prepaid or \$10 flash incentive
Nonresponse Conversion Phase (February 2021)		<ul style="list-style-type: none"> Incentive boost – \$10 boost incentive Mini survey for \$5 incentive (for B&B:16/17 respondents only toward the end of data collection) 	<ul style="list-style-type: none"> Abbreviated survey offered at \$30 As needed, mini survey for \$5 incentive (towards the end of data collection)
Total Incentives	\$5 Maximum=\$5	\$35 + \$2 prepaid \$5 early bird \$10 boost Maximum=\$47	\$30 \$2 prepaid or \$10 flash Maximum=\$32 or \$40

Note: The duration of each data collection phase will be based on whether phase capacity (see text above) has been reached; consequently, dates are mere estimates but could change depending on response rates.

Data Collection Protocol Design Elements

Greeting card. The first mailing that individuals in the aggressive and default data collection protocols will receive is a greeting card expressing gratitude for being part of the study and announcing the upcoming data collection, and in the aggressive protocol, the mailing of a \$2 prepaid incentive. Greeting cards have been shown to significantly increase response rates in longitudinal studies (Griggs et al. forthcoming) and we will use this method as a precursor to the invitation letter for both groups. The greeting card will be mailed a few weeks in advance of data collection upon OMB approval (anticipated by June 2020).

Eligibility screener. During the B&B:16/17 field test, 22% of the NPSAS:16 field test nonrespondents were determined ineligible by the B&B survey. For the B&B:16/20 main study, the non-study members and double nonrespondents will be sent an initial letter, email and text message inviting them to complete a brief *eligibility screener* online or by telephone (inbound calling only). Sample members who complete the eligibility screener will receive a \$5 promised incentive paid by their choice of check or PayPal. Requests to complete the screener will be mailed and emailed to sample members at the start of data collection, and more reminders will be sent during the

Early Completion Phase and the Production Phase I (see Table 3). Eligibility screening will begin with the start of data collection in July 2020.

Text messaging. Text message advance notifications and reminders have been shown to significantly increase response rates in different survey modes (e.g., Callegaro et al. 2011; Schober et al. 2015). Hard copy mailings, emails, and text messages will be used to maintain ongoing contact with sample members in all data collection protocols, prior to and throughout data collection. Text messaging will use RTI International’s Short Message Service (ARTEMIS). ARTEMIS is a network service built on top of an Adobe ColdFusion application server and Microsoft SQL Server data base engine to support research programs across a range of key areas, from simple messaging to data collection.

ARTEMIS texting will also be used to administer survey items on civic engagement. The main B&B:16/20 survey asks respondents if they are registered to vote in US elections and, for early respondents surveyed on or before November 3, 2020, asks whether they intend to vote in the November 2020 election; for those surveyed after November 3, 2020, the November voting question is revised to past tense. To collect information on November 3rd voting, a single follow-up item (“Did you vote in the November 2020 presidential election?”) will be fielded to early respondents via SMS text survey. The design of the civic engagement texted survey is presented in appendix F. Data collection is expected to occur from early November through early December.

Since this will be the first time texted surveying is employed with B&B populations, its success will be evaluated using a number of different factors. First, we will assess ease of use, based on both the effort required to program and send the survey and the number of attempts required to obtain a complete response. Second, observed participation rates will provide an early indicator of the viability of texting as a data collection mode. It is worth noting, however, that since it is only early completers who will be texted the additional survey item, participation rates may be higher than would be observed in the full sample. Finally, the data will be evaluated for completeness.

If the texted civic engagement survey is successful, we plan to offer it as an additional survey mode for the mini survey, described below. The mini survey is intended to collect a small set of critical items from all nonrespondents in the final phase of data collection. The texted mini survey will be offered only to those nonrespondents with known study eligibility. Limiting its availability is necessary because determining eligibility requires complex logic which cannot be implemented in the texted format. Nonrespondents with unknown eligibility will be able to complete the mini survey in web, including mobile, and CATI modes. A decision about inclusion of texting for the mini survey will be submitted in a change memorandum by February 2021.

Prepaid incentive. Cash prepaid incentives have been shown to significantly increase response rates in both interviewer-administered as well as self-administered surveys and hence reduce the potential for nonresponse bias (e.g., Church 1993; Cantor et al. 2008; Goeritz 2006; Medway and Tourangeau 2015; Messer and Dillman 2011; Parsons and Manierre 2014; Singer 2002). During the Early Completion Phase in the B&B:16/17 field test, prepaid incentives (\$10 via check or PayPal) in combination with telephone prompting also significantly increased response rates by 4.4 percentage points in the aggressive protocol implemented for prior round nonrespondents. Given these positive findings combined with general recommendations in the literature (e.g., Singer and Ye 2013), B&B:16/20 will send a

small prepaid incentive of \$2 in the data collection announcement letter to all sample members in the aggressive protocol (ever nonrespondents and B&B:16/17 abbreviated respondents - see also Calibration Experiment 1 discussion). Additionally, as a potential nonresponse conversion strategy, sample members in the default protocol (double respondents) may receive a prepaid incentive of \$2 as a final attempt to receive their full survey. This amount has been shown to effectively increase response rates at more efficient field costs compared to other prepaid incentives (e.g., Beebe et al. 2005; Millar and Dillman 2011; Tourangeau et al. 2013).

Results of the calibration experiment show that, while there is no statistically significant difference in response rates between the \$2 cash prepaid incentive and the \$2 PayPal prepaid incentive, there *is* a statistically significant difference in the resulting sample composition when it comes to age, as the Treatment Group results in a statistically significantly younger respondent sample. Therefore, individuals aged 30 and older will proceed with the incentive design from the calibration experiment control group (\$2 cash prepaid incentive) and individuals aged 29 and younger will proceed with the incentive design from the calibration experiment treatment group (\$2 PayPal prepaid incentive) for the B&B:16/20 main data collection in the aggressive protocol.

Sample members who receive a cash prepaid incentive and have “good” address information will receive the \$2 prepaid incentive in the mail. Sample members who are supposed to receive a cash prepaid incentive but for whom no good address information exists will receive the \$2 prepaid incentive via PayPal to their best-known e-mail address (e.g., in the B&B:16/17 full-scale cohort 47% of all respondents chose to receive their incentive via PayPal, and 46% of the B&B:08/18 full scale cohort). PayPal was successfully used for prepaid incentives in B&B:16/17 field test, B&B:08/18, and BPS:12/17. Once B&B:16/20 staff obtain good contacting information for a sample member, a \$2 cash incentive will be mailed out if the sample member has not yet claimed the \$2 PayPal offer and completed the survey (similar to the B&B:08/12 full-scale responsive design experiment). All data collection announcements related to interventions will be designed to stand out.

Early bird incentive. “Early bird” incentives have been shown to lead to faster responses and increased participation rates within the early bird incentive period (e.g., LeClere et al. 2012; Coppersmith et al. 2016), and can provide efficiencies by reducing both data collection costs and time. Given this, sample members in the aggressive protocol will be offered the opportunity to increase their total incentive by \$5 (for a total of \$42) when responding within the first three weeks of data collection. Sample members in the default protocol will not receive the early-bird incentive because, as shown in the B&B:16/17 full-scale or the B&B:08/18, they generally need less encouragement to participate.

Mode tailoring. The leverage-saliency theory suggests that respondents have different hooks that drive their likelihood of survey participation (Groves et al., 2000); thus, offering a person the mode they prefer may increase their likelihood of responding. This is further supported by empirical evidence that shows that offering people their preferred mode speeds up their response and is associated with higher participation rates (e.g., Olson et al. 2012). Using the B&B:16/17 survey completion mode as a proxy for mode preference, during the B&B:16/20 main study early completion phase sample members in the default protocol will be approached in the mode of completion for B&B:16/17. Specifically, while all sample members in the default protocol will receive identical data collection announcement letters and emails,

those who completed the B&B:16/17 survey by telephone (n=2,080; Wine et al. 2019) will be approached by telephone from the start of data collection. Likewise, those who completed the B&B:16/17 main study survey online will not be contacted by telephone before a preassigned outbound telephone data collection date.

Light outbound CATI calling. Light CATI involves a minimal number of phone calls, used mainly to prompt web response (as opposed to regular CATI efforts that involve more frequent phone efforts, with the goal to locate sample members and encourage their participation). In the B&B:16/17 field test, introduction of light CATI interviewing appeared to increase production phase response rates in the default protocol. Although one should use caution when interpreting these results – group assignment in B&B:16/17 field test was not random but instead compared NPSAS:16 ‘early’ and ‘late’ respondents– the findings are consistent with the literature which has shown that web surveys tend to have lower response rates compared to interviewer-administered surveys (e.g., Lozar, Manfreda et al. 2008). Attempting to survey sample members by telephone also increases the likelihood of initiating locating efforts sooner. B&B:16/17 field test results showed higher locate rates in the default protocol (93.7%), which had light CATI, compared to a more relaxed protocol without light CATI (77.8%; $\chi^2 = 63.2$, $p < 0.001$). For the B&B:16/20 main study data collection, light CATI will be used with the default protocol once CATI begins in Production Phase I.

Abbreviated survey. Obtaining responses from all sample members in a data collection is an important assumption of the inferential paradigm. The leverage-saliency theory and the social exchange theory suggest that the participation decision of an individual is driven by different survey design factors or perceived cost of participating. As such, reducing the perceived burden of participating by reducing the survey length may motivate sample members to participate.

During the B&B:16/17 field test data collection, sample members in the aggressive protocol (prior round nonrespondents) who were offered the abbreviated survey during the production phase responded at higher rates (22.7%) than those who were not offered the abbreviated survey at the same time (12.1%; $t(2,097) = 3.67$, $p < 0.001$). The B&B:08/12 and B&B:08/18 full-scale studies showed similar results in that offering the abbreviated survey to prior round nonrespondents at a later point in data collection increased overall response rates of that group by 18.2 and 8.8 percentage points respectively (Cominole et al. 2015). An abbreviated survey option will be offered to all sample members in the B&B:16/20 main study data collection. For the aggressive protocol, the abbreviated survey will be offered during Production Phase II, which is the latter half of the production phase of data collection, and as the last step in nonresponse conversion in the default protocol.

\$2 Prepaid or Flash Incentive for Nonresponse Conversion. Incentive boosts are successful nonresponse conversion strategies, increasing response rates across various modes of data collection (e.g., Singer and Ye, 2013; Dykema et al., 2015; Stevenson et al., 2016; Lynn, 2017). Furthermore, evidence from surveys in different modes suggests that *prepaid* incentives are more effective than promised incentives in increasing response rates (e.g., Singer et al. 1999; Singer, 2002; Mercer et al. 2015). However, little is known of the effect of prepaid incentives at *later* phases of data collection; thus, we recommend exploring whether offering a lower value *prepaid* incentive may outweigh the benefit of offering a higher *promised* incentive. If response rates warrant, we recommend implementing an experiment in the default data collection protocol to test these two incentive strategies targeted at nonresponse conversion as a final attempt to obtain a full survey. If implemented, this experiment

will provide insights regarding optimal nonresponse conversion strategies in future B&B studies.

More specifically, we propose to compare the effectiveness of a \$2 prepaid incentive, sent with a reminder letter, to that of a \$10 promised flash incentive⁵ which will temporarily increase the baseline incentive from \$30 to \$40 if the full survey is completed within two weeks of the reminder. This experiment is conditional on the response rate achieved toward the end of Production Phase II and will be utilized if the response rate is 70% or lower (see discussion below):

- **Treatment Group 1** will receive a \$2 prepaid incentive with a reminder letter.
- **Treatment Group 2** will receive a \$10 promised flash incentive increase if they complete the full survey within the two weeks following the reminder.

We recommend a disproportionate assignment of sample members to treatment groups: towards the end of Production Phase II we will randomly assign 33% of the remaining prior-round double respondents to the Treatment Group 1 and the remainder to the Treatment Group 2.

This experiment will only be implemented if there is enough sample size to yield sufficient power for a 2-group comparison (see discussion below). There are two possible scenarios that depend on response rates obtained during earlier data collection phases:

- 1) The response rate is above 70%: Due to a sample size that is too small to merit an experimental comparison, we recommend implementing the \$10 flash incentive without any experimentation, since previous B&B implementations of a flash incentive have shown to successfully increase response rate.
- 2) The response rate is 70% or lower: There is enough sample size to implement an experiment.

This approach would allow for comparisons of response rates among two experimental groups with at least 1,488 sample members in Treatment Group 1 and at least 3,021 sample members in Treatment Group 2 and provide enough power to detect at least a 5-percentage point difference in response rates assuming 80 percent power, type I error of 5 percent and a base response rate of 50 percent among the remaining nonrespondents. This calculation assumes a 2-sided chi-square test of the response proportions.

The experiment described above will allow us to test the following hypotheses:

- H2.1. There is no statistically significant difference in response rates between Treatment Group 1 and Treatment Group 2.
- H2.2. There is no statistically significant difference in representativeness (demographic characteristics) between Treatment Group 1 and Treatment Group 2.
- H2.3. There is no statistically significant difference in level of effort between Treatment Group 1 and Treatment Group 2.

The proposed experimental period is expected to run for 2 weeks at the end Production Phase II (likely in January 2021). This particular timing ensures that the “flash” intervention occurs towards the end of the full survey collection phase just

⁵ Offering the flash incentive increased response rates among the B&B:08/18 full-scale double respondents by 2.2 percentage points.

before the offer to complete the abbreviated survey is mailed out. This timing also ensures that most sample members have been located. Results will be submitted to OMB via change memorandum by February 2021.

Incentive boosts. Researchers have commonly used incentive boosts as a nonresponse conversion strategy for sample members who have implicitly or explicitly refused to complete the survey (e.g., Groves and Heeringa 2006; Singer and Ye 2013). These boosts are especially common in large federal surveys during their nonresponse follow-up phase (e.g., The National Survey of Family Growth) and have been implemented successfully in other postsecondary education surveys (e.g., HSLs:09 F2; BPS:12/17). For nonresponse conversion, a \$10 incentive boost increase to the B&B:16/20 baseline incentive is planned for all remaining nonrespondents in the aggressive protocol, approximately three months after the abbreviated survey offer and about four weeks before the end of data collection. In the B&B:08/18 full scale, offering an incentive boost to prior round nonrespondents increased response rates to the abbreviated survey by 5.6 percentage points.

Mini survey. Obtaining information on the critical survey items for nonrespondents is crucial to assess the potential for nonresponse bias and better inform imputation. The shorter the stated length of a survey, the lower the perceived burden for the respondent. Motivated by this approach, B&B:16/20 will offer an extremely abbreviated questionnaire – a mini survey – in the aggressive protocol to all B&B:16/17 respondents who have not completed B&B:16/20 late in the data collection period and to sample members in the default protocol if we have not reached phase capacity with the abbreviated survey.

The mini survey, presented in Appendix E, contains only the most critical survey items and is estimated to take about 5 minutes. Similarly, to the NPSAS:20 mini survey, this request will be associated with a \$5 promised incentive. By offering the mini survey, we expect to further increase response rates compared to the traditional abbreviated interview given the smaller burden of the request. As described above, conditional on the success of the text message survey about civic engagement, we will consider the option of fielding the mini survey as a text message survey to sample members in the default protocol.

Other interventions. While all B&B studies are conducted by NCES, the data collection contractor, RTI International, has typically used a study-specific e-mail “@rti.org” and telephone number to contact and support sample members. For the B&B:08/18 field test, B&B:08/18 staff investigated the effect of sending reminder e-mails from an “@ed.gov” e-mail address. Compared to sending e-mails from “@rti.org,” the B&B:08/18 field test showed that sending reminders from NCES had positive effects on sample representativeness and data collection efficiency (see B&B:08/18 OMB # 1850-0729 v. 13 Appendix C). B&B:16/20 full-scale e-mails will be sent from “@ed.gov” with occasional e-mail reminders sent either from the NCES project officer or the RTI project director. Changing the e-mail sender to the project officer may increase the perceived importance of the survey and help personalize the contact materials, thereby potentially increasing relevance. Switching the sender during data collection also increases the chance that the survey invitation is delivered to the sample member rather than to a spam filter.

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

The study is being conducted by NCES. The following statisticians at NCES are responsible for the statistical aspects of the study: Mr. Ted Socha, Dr. Tracy Hunt-White, Dr. David Richards, and Dr. Gail Mulligan. NCES's prime contractor for B&B:16/20 is RTI International (RTI). The following staff members at RTI are working on the statistical aspects of the study design: Dr. Jennifer Wine, Dr. Melissa Cominole, Ms. Jennifer Cooney, Dr. Nicole Tate, Ms. Erin Thomsen, Dr. Antje Kirchner, Dr. Erin Dunlop Velez, Dr. Emilia Peytcheva, and Mr. Peter Siegel.

Subcontractors include HR Directions; Research Support Services; EurekaFacts, LLC; ManTech International; Activate Research; and Strategic Communications, Inc. The consultants are Dr. Sandy Baum, Dr. Stephen Porter, and Dr. Paul D. Umbach. Principal professional RTI staff, not listed above, who are assigned to the study include Ms. Donna Anderson, Ms. Gayathri Bhat, and Ms. Ashley Wilson.

C. References

- American Association for Public Opinion Research. 2016. Standard Definitions Final Dispositions of Case Codes and Outcome Rates for Surveys. Retrieved 05/07/2020: https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf.
- Beebe, T.J., Davern, M.E., McAlpine, D.D., Call, K.T., and Rockwood, T.H. (2005). Increasing response rates in a survey of Medicaid enrollees: The effect of a prepaid monetary incentive and mixed modes (mail and telephone). *Medical Care*, 43, 411- 420.
- Callegaro, M. Ayhan, O., Gabler, S., Haeder, S., & Villar, A. (2011). Combining landline and mobile phone samples. A dual frame approach. Working Papers 2011/13. Gesis Leibniz-Institut fuer Sozialwissenschaften.
- Cantor, D., O'Hare, B.C., and O'Connor, K.S. (2008). The Use of Monetary Incentives to Reduce Nonresponse in Random Digit Dial Telephone Surveys. In Lepkowski, J.M., Tucker, N.C., Brick, J.M., de Leeuw, E., Japec, L., Lavrakas, P.J., Link, M.W., and Sangster, R.L. (eds.). *Advances in Telephone Survey Methodology*. New York: Wiley.
- Church, A.H. (1993). Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis. *Public Opinion Quarterly*, 57(1), 62-79.
- Cominole, M., Shepherd, B., and Siegel, P. (2015). 2008/12 Baccalaureate and Beyond Longitudinal Study (B&B:08/12) Data File Documentation (NCES 2015-141). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved [1/22/1018] from <http://nces.ed.gov/pubsearch>.
- Coppersmith, J., Vogel, L.K., Bruursema, T., and K. Feeney. (2016). Effects of Incentive Amount and Type of Web Survey Response Rates. *Survey Practice*, no pp.
- DeBell, M., Maisel, N., Edwards, B., Amsbary, M., and Meldener, V. (2019). Improving Survey Response Rates with Visible Money. *Journal of Survey Statistics and Methodology*. Online First. Retrieved from: <https://academic.oup.com/jssam/advance-article/doi/10.1093/jssam/smz038/5610622>
- Dykema, J., Jaques, K., Cyffka, K., Assad, N., Hammers, R.G., Elver, K., Meliecki, K.C., and Stevenson, J. (2015). Effects of Sequential Prepaid Incentives and Envelope Messaging in Mail Surveys. *Public Opinion Quarterly*, 79(4), 906-931.
- Goeritz, A.S. (2006). Incentives in web studies: Methodological issues and review. *International Journal of Internet Science*, 1, 58-70.
- Griggs, A., Powell, R., Keeney, J., Waggy, M., Harris, K., Halpern, C. and Dean, S. (forthcoming) Research Note: A Prenotice Greeting Card's Impact on Response Rates and Response Time, *Longitudinal and Life Course Studies*.
- Groves R.M., and Heeringa, S.G. (2006). Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs. *Journal of the Royal Statistical Society Series A-Statistics in Society*, 169(3), 439-457.
- Groves, R.M., Singer, E. and Corning, A. (2000). Leverage-Saliency Theory of Survey Participation. Description and Illustration. *Public Opinion Quarterly*, 64, 299-308.

- LeClere, F., Plumme, S., Vanicek, J., Amaya, A., and K. Carris. (2012). Household Early Bird Incentives: Leveraging Family Influence to Improve Household Response Rates." American Statistical Association Joint Statistical Meetings, Section on Survey Research.
- Lozar Manfreda, K., Bosnjak, M., Berzelak, J., Haas, I., and Vehovar, V. (2008). Web Surveys Versus Other Survey Modes. A Meta-Analysis Comparing Response Rates. *International Journal of Market Research*, 50(1), 79-104.
- Lynn, P. (2017). From Standardised to Targeted Survey Procedures for Tackling Non-response and Attrition. *Survey Research Methods*, 11(1), 93-103.
- 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.
- Mercer, A., Caporaso, A., Cantor, D., and Townsend, R. (2015). How Much Gets You How Much? Monetary Incentives and Response Rates in Household Surveys. *Public Opinion Quarterly*, 79(1), 105-129.
- Millar, M.M., and Dillman, D.A. (2011). Improving Response to Web and Mixed-Mode Surveys. *Public Opinion Quarterly*, 75(2), 249-269.
- Olson, K., Smyth, J. D., and Wood, H. (2012). Does Giving People Their Preferred Survey Mode Actually Increase Survey Participation? An Experimental Examination. *Public Opinion Quarterly*, 76, 611-635.
- Parsons, L., and Manierre, M.J. (2014). Investigating the Relationship among Prepaid Token Incentives, Response Rates, and Nonresponse Bias in a Web Survey. *Field Methods*, 26(2), 191-204.
- Schober, M.F., Conrad, F.G., Antoun, C., Ehlen, P., Fail, S., Hupp, A.L., Johnston, M., Vickers, L., Yan, Y.H., and Zhang, C. (2015). Precision and Disclosure in Text and Voice Interviews on Smartphones. *PLoS ONE* 10(6): e0128337. doi:10.1371/journal.pone.0128337
- Singer, E. (2002). The Use of Incentives to Reduce Nonresponse in Household Surveys. In Groves, R.M., Dillman, D. A., Eltinge, J.L., Little, R.J.A. (eds.), *Survey Nonresponse*. New York: Wiley.
- Singer, E. and Ye, C. (2013). The Use and Effects of Incentives in Surveys. *Annals of the American Academy of Political and Social Science*, 645(1), 112-141.
- Stevenson, J., Dykema, J., Kniss, C., Assad, N., and Taylor, C. (2016). Effects of Sequential Prepaid Incentives to Increase Participation and Data Quality in a Mail Survey of Pediatricians. Paper Presented at the Annual Conference of the Midwest Association for Public Opinion Research, Nov.
- Tourangeau, R., Conrad, F.G., and Couper, M. (2013). *The Science of Web Surveys*. Oxford, NY: Oxford University Press.
- Wine, J., Tate, N., Thomsen, E., Cooney, J., Haynes, H. (2020) 2016/17 Baccalaureate and Beyond Longitudinal Study 51 (B&B:16/17) Data File Documentation (NCES 2020-441). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Retrieved [09/26/2019] from <https://nces.ed.gov/pubsearch>.

