

2012/17 Beginning Postsecondary Students Longitudinal Study: (BPS:12/17)

Full Scale Interview and Administrative Record
Collections

Supporting Statement Part B
OMB # 1850-0631 v.10

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

August 2016

Contents

B. Collection of Information Employing Statistical Methods.....	3
1. Respondent Universe.....	3
a. NPSAS:12 Full-scale Institution Universe.....	3
b. NPSAS:12 Full-scale Student Universe.....	3
2. Statistical Methodology.....	4
a. NPSAS:12 Full-scale Sample.....	4
b. BPS:12/14 Full-scale Sample.....	7
c. BPS:12/17 Full-scale Sample.....	7
d. BPS:12 Student Records and PETS Samples.....	8
3. Methods for Maximizing Response Rates.....	9
a. Locating.....	10
b. Prompting Cases with Mail, Email, and SMS Contacts.....	10
c. Telephone Interviewing and Help Desk Support.....	11
d. Quality Control.....	12
e. Administrative Record Collections.....	13
4. Tests of Procedures and Methods.....	17
a. BPS:12/14 Full Scale.....	17
b. BPS:12/17 Full Scale.....	22
c. BPS:12 PETS Pilot Collection.....	31
5. Reviewing Statisticians and Individuals Responsible for Designing and Conducting the Study.....	32
C. References.....	32

Tables

Table 1. BPS:12/14 sample size by institution characteristics: 2012.....	7
Table 2. BPS:12/17 sample member disposition, by NPSAS:12 and BPS:12/14 response status.....	8
Table 3. BPS:12/17 sample size by institution characteristics: 2016.....	8
Table 4. Pilot sample member disposition, by sample member and prior round response status.....	9
Table 5. Summary of start dates and activities for each phase of the BPS:12/14 data collection, by sample: 2014.....	19
Table 6. Summary of responsive design impact on nonresponse bias, by institutional sector: 2014.....	21
Table 7. Candidate variables for the bias likelihood model.....	24
Table 8. Targeted sector groups.....	25
Table 9. Timeline for previous respondents.....	26
Table 10. Timeline for NPSAS:12 study member double interview nonrespondents.....	27
Table 11. Power for control versus treatment comparisons across multiple assumptions.....	30
Table 12. Power to detect 5 percent difference in response rate based on different baseline response rates.....	31

B. Collection of Information Employing Statistical Methods

This section describes the target population for the 2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17) and the sampling and statistical methodologies proposed for the second follow-up full- scale study. This section also addresses suggested methods for maximizing response rates and for tests of procedures and methods, and introduces the statisticians and other technical staff responsible for design and administration of the study.

1. Respondent Universe

The target population for the 2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17) consists of all students who began their postsecondary education for the first time during the 2011–12 academic year at any Title IV-eligible postsecondary institution in the United States. The sample students were the first-time beginning (FTB) students from the 2011-12 National Postsecondary Student Aid Study (NPSAS:12).

a. NPSAS:12 Full-scale Institution Universe

To be eligible for NPSAS:12, students must have been enrolled in a NPSAS-eligible institution in any term or course of instruction at any time during the 2011–12 academic year. Institutions must have also met the following requirements:

- offer an educational program designed for persons who have completed secondary education;
- offer at least one academic, occupational, or vocational program of study lasting at least 3 months or 300 clock hours;
- offer courses that were open to more than the employees or members of the company or group (e.g. union) that administers the institution;
- be located in the 50 states or the District of Columbia;
- not be a U.S. service academy institution; and
- have signed the Title IV participation agreement with ED.¹

NPSAS excluded institutions providing only avocational, recreational, or remedial courses, or only in-house courses for their own employees or members. U.S. service academies were also excluded because of the academies' unique funding/tuition base.

b. NPSAS:12 Full-scale Student Universe

Students eligible for the NPSAS:12 full-scale were those who attended a NPSAS eligible institution during the 2011–12 academic year and who were

- enrolled in *either*: (a) an academic program; (b) at least one course for credit that could be applied toward fulfilling the requirements for an academic degree; (c) exclusively non-credit remedial coursework but determined by the institution to be eligible for Title IV aid; or (d) an occupational or vocational program that required at least 3 months or 300 clock hours of instruction to receive a degree, certificate, or other formal award;
- not currently enrolled in high school; and
- not solely enrolled in a General Educational Development (GED) or other high school completion program.

¹ A Title IV eligible institution is an institution that has a written agreement (program participation agreement) with the U.S. Secretary of Education that allows the institution to participate in any of the Title IV federal student financial assistance programs other than the State Student Incentive Grant (SSIG) and the National Early Intervention Scholarship and Partnership (NEISP) programs.

The full-scale student sampling frame was created from lists of students enrolled at NPSAS sampled institutions between July 1, 2011 and April 30, 2012. While the NPSAS:12 study year covers the time period between July 1, 2011 and June 30, 2012 to coincide with the federal financial aid award year, the abbreviated sampling period of July 1 to April 30 facilitated timely completion of data collection and data file preparation. Previous cycles of NPSAS have shown that the terms beginning in May and June add little to enrollment and aid totals.

To create the student sampling frame, each participating institution was asked to submit a list of eligible students with the following data provided for each listed student:

- name;
- student ID (if different than Social Security number);
- Social Security number;
- date of birth;
- date of high school graduation (month and year);
- degree level during the last term of enrollment (undergraduate, masters, doctoral-research/scholarship/other, doctoral-professional practice, or other graduate);
- class level if undergraduate (first, second, third, fourth, or fifth year or higher);
- undergraduate degree program;
- Classification of Instructional Program code or major;
- FTB status; and
- contact information.

Requesting contact information for eligible students prior to sampling allowed for student interviewing to begin shortly after sample selection, which facilitated management of the schedule for data collection, data processing, and file development.

2. Statistical Methodology

Because the students in the BPS:12/17 full-scale sample come from the NPSAS:12 full-scale sample, this section describes the NPSAS:12 full-scale sample design, which was a two-stage sample consisting of a sample of institutions at the first stage, and a sample of students from within sampled institutions at the second stage. The BPS:12/17 full-scale sample comprised students from the NPSAS:12 full-scale sample who were determined to be FTBs, or were identified by the NPSAS institution as potential FTBs. Because the BPS:12/17 sample corresponds to eligible sample members from the first follow-up with BPS:12 cohort, BPS:12/14, this section also presents information about the BPS:12/14 full-scale sample design.

a. NPSAS:12 Full-scale Sample

The initial institution samples for the NPSAS field test and full-scale studies were selected simultaneously, prior to the full-scale study, using sequential probability minimum replacement (PMR) sampling (Chromy 1979), which resembles stratified systematic sampling with probabilities proportional to a composite measure of size (Folsom, Potter, and Williams 1987). This is the same methodology that has been used since NPSAS:96. Institution measure of size was determined using annual enrollment data from the most recent IPEDS 12-Month Enrollment Component and first time beginner (FTB) full-time enrollment data from the most recent IPEDS Fall Enrollment Component. Composite measure of size sampling was used to ensure that target sample sizes were achieved within institution and student sampling strata, while also achieving approximately equal student weights across institutions. The institution sampling frame for NPSAS:12 full-scale was constructed using the 2009 Integrated Postsecondary Education Data System (IPEDS) header, Institution Characteristics (IC), Fall and 12-

Month Enrollment, and Completions files. All eligible students from sampled institutions composed the student sampling frame.

From the stratified frame, a total of 1,970 institutions were selected to participate in either the field test or full-scale study. Using simple random sampling within institution strata, a subsample of 300 institutions was selected for the field test sample, with the remaining 1,670 institutions comprising the sample for the full-scale study. This sampling process eliminated the possibility that an institution would be burdened with participation in both the field test and full-scale samples, and ensured representativeness of the full-scale sample.

The institution strata used for the sampling design were based on institution level, control, and highest level of offering, including:

1. public less-than-2-year,
2. public 2-year,
3. public 4-year non-doctorate-granting,
4. public 4-year doctorate-granting,
5. private nonprofit less-than-4-year,
6. private nonprofit 4-year non-doctorate-granting,
7. private nonprofit 4-year doctorate-granting,
8. private for-profit less-than-2-year,
9. private for-profit 2-year, and
10. private for-profit 4-year.

Due to the growth of the for-profit sector, private for-profit 4-year and private for-profit 2-year institutions were separated into their own strata, unlike in previous administrations of NPSAS.

In order to approximate proportional representation of institutions within each institution stratum, additional implicit stratification for the full-scale was accomplished by sorting the sampling frame by the following classifications: (1) historically Black colleges and universities indicator; (2) Hispanic-serving institutions indicator; (3) Carnegie classifications of degree-granting postsecondary institutions; (4) 2-digit Classification of Instructional Programs (CIP) code of the largest program for less-than-2-year institutions; (5) the Office of Business Economics Region from the IPEDS header file (Bureau of Economic Analysis of the U.S. Department of Commerce Region); (6) state and system for states with large systems, e.g., the SUNY and CUNY systems in New York, the state and technical colleges in Georgia, and the California State University and University of California systems in California; and (7) the institution measure of size.

The institution sample was freshened in order to add newly eligible institutions to the sample. The newly-available 2009–10 IPEDS IC header, 12-month and fall enrollment, and completions files were used to create an updated sampling frame of current NPSAS-eligible institutions. This frame was compared to the original frame and 387 new or newly eligible institutions were identified for the freshened sampling frame. A sample size of 20 was selected in order to produce similar probabilities of selection to the originally selected institutions within sector (stratum) in order to minimize unequal weights and subsequently variances.

From the 1,690 institutions selected for the NPSAS:12 full-scale data collection, 100 percent met eligibility requirements; of those, approximately 88 percent (or 1,480 institutions) provided enrollment lists. The NPSAS:12 full-scale sample was randomly selected from the frame with students sampled at fixed rates according to student education level and institution sampling strata. Sample yield was monitored and sampling rates were adjusted when necessary. The full-scale sample achieved a size of about 128,120 students, of which approximately 59,740 were potential FTB students, 51,050 were other

undergraduate students, and 17,330 were graduate students. The achieved sample size was higher than originally targeted because institution participation rates were higher than estimated, sampling continued longer than scheduled, and a higher sample size was desired to help meet interview yield targets.

Identification of FTB students in NPSAS:12. Close attention was paid to accurately identify FTB students in NPSAS to avoid unacceptably high rates of misclassification (e.g., false positives)² which, in prior BPS administrations, have resulted in (1) excessive cohort loss, (2) excessive cost to “replenish” the sample, and (3) an inefficient sample design (excessive oversampling of “potential” FTB students) to compensate for anticipated misclassification errors. To address this concern, participating institutions were asked to provide additional information for all eligible students, and matching to administrative databases was utilized to further reduce false positives prior to sample selection.

Participating institutions were asked to provide the FTB status and high school graduation date for every listed student. High school graduation date was used to remove students from the frame who were co-enrolled in high school. FTB status, along with class and student levels, were used to exclude misclassified FTB students in their third year or higher and/or those who were not an undergraduate student. FTB status, along with date of birth, were used to identify students older than 18 who were sent for pre-sampling matching to administrative databases in an effort to confirm FTB status.

If the FTB indicator was not provided for a student on the lists, but the student was 18 years of age or younger and did not appear to be dually enrolled in high school, the student was sampled as an FTB student. Otherwise, if the FTB indicator was not provided for a student on the list and the student was over the age of 18, then the student was sampled as an “other undergraduate” (but such students would be included in the BPS cohort if identified during the student interview as an FTB student).

Prior to sampling, students over the age of 18 listed as potential FTB students were matched to NSLDS records to determine if any had a federal financial aid history pre-dating the NPSAS year (earlier than July 1, 2011). Since NSLDS maintains current records of all Title IV federal grant and loan funding, any student with disbursements from the prior year or earlier could be reliably excluded from the sampling frame of FTB students. Given that about 60 percent of FTB students receive some form of Title IV aid in their first year, this matching process could not exclude all listed FTB students with prior enrollment, but significantly improved the accuracy of the list prior to sampling, yielding fewer false positives. After undergoing NSLDS matching, students over the age of 18 still listed as potential FTB students were matched to the National Student Clearinghouse (NSC) for further narrowing of potential FTB students based on evidence of earlier enrollment.

Matching to NSLDS identified about 20 percent of cases as false positives and matching to NSC identified about 7 percent of cases as false positives. In addition to NSLDS and NSC, a subset of potential FTB students on the student sampling frame was sent to the Central Processing System (CPS) for matching to evaluate the benefit of the CPS match for the full-scale study. Of the 2,103,620 students sent, CPS identified about 17 percent as false positives. Overall, matching to all sources identified about 27 percent of potential FTB students over the age of 18 as false positives, with many of the false positives identified by CPS also identified by NSLDS or NSC. The matching appeared most effective in identifying false positives (i.e., non-FTB students) among public 2-year and private for-profit institutions. While public less-than 2-year and private nonprofit less-than-4-year institutions have a high percent of false positives, they represent a small percentage of the total sample.

Since this pre-sampling matching was new to NPSAS:12, the FTB sample size was set high to ensure that a sufficient number of true FTB students would be interviewed. In addition, FTB selection rates were set taking into account the error rates observed in NPSAS:04 and BPS:04/06 within each sector. Additional information on NPSAS:04 methodology is available in that study’s methodology report (publication number NCES 2006180), or the BPS:04/06 methodology report (publication number NCES 2008184).

² A student identified by the institution on the enrollment list as an FTB who turns out to not be an FTB is a false positive.

These rates were adjusted to reflect the improvement in the accuracy of the frame from the NSLDS and NSC record matching. Sector-level FTB error rates from the field test were used to help determine the rates necessary for full-scale student sampling.

b. BPS:12/14 Full-scale Sample

At the conclusion of the NPSAS:12 full-scale, 30,076 students had been interviewed and confirmed to be FTB students, and all were included in the BPS:12/14 full-scale. In addition, the full-scale sample included the 7,090 students who did not respond to the NPSAS:12 full-scale, but were potential FTB students according to student records or institution lists. The distribution of the BPS:12/14 full-scale sample is shown in table 1, by institution sector.

Table 1. BPS:12/14 sample size by institution characteristics: 2012

Institution characteristics	Total	Confirmed and potential FTB students from the NPSAS:12 full-scale		
		NPSAS interview respondent	NPSAS study member interview nonrespondents	NPSAS non-study member interview nonrespondents
Total	37,170	30,080	4,610	2,480
Institution type				
Public				
Less-than-2-year	250	190	30	30
2-year	11,430	9,080	1,290	1,050
4-year non-doctorate-granting	1,930	1,700	150	90
4-year doctorate-granting	3,510	3,220	220	80
Private nonprofit				
Less-than-4-year	380	310	40	30
4-year non-doctorate-granting	2,430	2,160	120	140
4-year doctorate-granting	2,720	2,470	130	120
Private for-profit				
Less-than-2-year	1,630	1,200	380	50
2-year	3,530	2,620	710	200
4-year	9,370	7,110	1,550	700

NOTE: Detail may not sum to totals because of rounding. Potential FTB students are those NPSAS:12 non-respondents who appeared to be FTB students in NPSAS:12 student records data. Non study members are those NPSAS:12 sample members who, across all data sources, did not have sufficient data to support the analytic objectives of the study.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2011–12 National Postsecondary Student Aid Study (NPSAS:12) Full-scale.

c. BPS:12/17 Full-scale Sample

The BPS:12/17 sample is a subset of the BPS:12/14 sample in that BPS:12/14 sample members determined to not be FTB students are ineligible and, therefore, excluded from the BPS:12/17 sample. Furthermore, deceased individuals are also excluded from the BPS:12/17 sample. Table 2 provides the distribution of the 35,540 eligible BPS:12/17 sample members, by NPSAS:12 and BPS:12/14 response status.

Table 2. BPS:12/17 sample member disposition, by NPSAS:12 and BPS:12/14 response status.

Study member status and interview response status	Number of eligible cases	Field in BPS:12/17
Total	35,540	
NPSAS:12 study member		
NPSAS:12 interview respondent		
(1) BPS:12/14 respondent	23,640	Yes
(2) BPS:12/14 nonrespondent	5,720	Yes
NPSAS:12 interview nonrespondent		
(3) BPS:12/14 respondent	800	Yes
(4) BPS:12/14 nonrespondent	3,280	Yes
NPSAS:12 non-study member		
NPSAS:12 interview respondent		
(5) BPS:12/14 respondent	20	Yes
(6) BPS:12/14 nonrespondent	5	No
NPSAS:12 interview nonrespondent		
(7) BPS:12/14 respondent	320	Yes
(8) BPS:12/14 nonrespondent	1,785	No

NOTE: Detail may not sum to totals because of rounding. The total does not include 1,630 cases determined to be study ineligible or deceased as of BPS:12/14. Non-study members are those NPSAS:12 sample members who, across all data sources, did not have sufficient data to support the analytic objectives of the study.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17) Full-scale.

While the 1,790 sample members who did not respond to the BPS:12/14 interview and lacked sufficient information to be classified as NPSAS:12 study members are eligible for BPS:12/17, these sample members (groups 6 and 8 in table 2) will not be fielded for BPS:12/17. Instead, they will be treated as study nonrespondents for purposes of response rate calculation and will be accounted for with weight adjustments. However, for BPS:12 Student Records and PETS, records for all eligible sample members will be collected. The distribution of the remaining 33,750 sample members across institutional sector is provided in table 3.

Table 3. BPS:12/17 sample size by institution characteristics: 2016

Institution characteristics	Total
Total	33,750
Institution type	
Public	
Less-than-2-year	210
2-year	10,150
4-year non-doctorate-granting	1,830
4-year doctorate-granting	3,400
Private nonprofit	
Less-than-4-year	330
4-year non-doctorate-granting	2,280
4-year doctorate-granting	2,600
Private for-profit	
Less-than-2-year	1,470
2-year	3,140
4-year	8,350

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17) Full-scale.

d. BPS:12 Student Records and PETS Samples

Student Records Sample. Student records will be collected from all known institutions of all eligible sample members.

PETS Pilot Test Sample. Transcripts will be requested for 1,167 sample members who responded to the BPS:12/17 pilot test interview conducted in spring 2016. This collection of transcripts is for methodological analyses, as described in section B.4.c. The resulting data will not be released, but the results will be published in the data file documentation.

The BPS:12/17 pilot test interview respondents were part of the sample that began with the NPSAS:12 field test. At the conclusion of the NPSAS:12 field test, approximately 2,000 students had been interviewed and confirmed to be FTB students. The BPS:12/14 field test included all of the students who responded to the NPSAS:12 field test and were confirmed to be FTBs, plus approximately 1,500 students who did not respond to the NPSAS:12 field test but were potential FTBs according to institution lists and/or student records. As shown in table 4, of the 3,496 field-test sample members, 143 were not FTBs and, therefore, were ineligible (98 NPSAS:12 study members and 45 NPSAS:12 non-study members). Of the remaining 3,353 eligible sample members, 1,884 responded to the BPS:12/14 field test.

For the BPS:12/17 pilot test, sample members who did not meet the definition of a NPSAS:12 study member were excluded, as well as sample members who did not respond to NPSAS:12 and were either deemed ineligible in, or did not respond to, BPS:12/14.

Table 4. Pilot sample member disposition, by sample member and prior round response status

Study member and prior-round response status	Number in pilot sample	Number fielded in BPS:12/17 pilot
Total	3,496	2,308
NPSAS:12 non-study member	347	0
NPSAS:12 nonrespondent	343	0
BPS:12/14 ineligible	45	0
BPS:12/14 nonrespondent	233	0
BPS:12/14 respondent	65	0

NPSAS:12 respondent	4	0
BPS:12/14 nonrespondent	4	0
NPSAS:12 study member	3,149	2,308
NPSAS:12 nonrespondent	1,150	309
BPS:12/14 ineligible	98	0
BPS:12/14 nonrespondent	743	0
BPS:12/14 respondent	309	309
NPSAS:12 respondent	1,999	1,999
BPS:12/14 nonrespondent	489	489
BPS:12/14 respondent	1,510	1,510

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17), Pilot Test.

Of the 2,308 sample members fielded in BPS:12/17 pilot study, 1,167 responded. The transcripts of these 1,167 sample members will be requested from known postsecondary institutions.

PETS Full-Scale Sample. Transcripts will be collected from all known institutions of all eligible sample members.

3. Methods for Maximizing Response Rates

Methods for maximizing response rates will be applied to the student interview, student records, and transcript activities independently. Maximizing student interview response rates will focus on locating sample members, prompting them to complete the interview, and offering convenient interview modes, such as web-based and telephone interviews. Maximizing administrative data collections of student records and transcripts will focus on gaining cooperation of institution representatives through mail, email, and telephone contacts, minimizing burden by offering a variety of means of providing data, and offering guidance and assistance when needed.

a. Locating

Achieving a high response rate for BPS:12/17 depends on successfully locating sample members and gaining their cooperation. The availability, completeness, and accuracy of the locating data collected in the NPSAS:12 and BPS12/14 data collections will impact the success of locating efforts for BPS:12/17. For BPS:12/17, all contact information previously collected for sample members, including current and prior address information, telephone numbers, and email addresses, will be stored in a centralized locator database. This database provides telephone interviewers and tracers with immediate access to all contact information available for BPS:12/17 sample members and to new leads developed through locating efforts during data collection.

BPS locating procedures use a multi-tiered tracing approach, in which the most cost-effective steps will be taken first to minimize the number of cases that require more expensive tracing efforts:

1. **Advance Tracing** primarily consists of national database batch searches conducted prior to the start of data collection. This step capitalizes on the locating data collected during NPSAS:12 and BPS:12/14.
2. **Telephone Locating and Interviewing** includes calling all telephone leads, prompting sample members to complete the online or telephone interview, attempting to get updated contact information from parents or other contacts, and following-up on leads generated by these contact attempts.

3. **Pre-Intensive Batch Tracing** is conducted between telephone locating efforts and intensive tracing, and is intended to locate cases as inexpensively as possible before moving on to more costly intensive tracing efforts.
4. **Intensive Tracing** consists of tracers checking all telephone numbers and conducting credit bureau database searches after all current telephone numbers have been exhausted.
5. **Other Locating Activities** will take place as needed and may include use of social networking sites and use of new tracing resources.

Initial Contact Update. BPS:12/17 will conduct an initial contact address update to encourage sample members to update their contact information prior to the start of data collection. The initial contact mailing and email will introduce the study and ask sample members to update their contact information via the study website. The address update will be conducted a few weeks prior to the start of full-scale data collection.

In addition, in the fall of 2016, we will conduct panel maintenance activities on full-scale sample members to encourage them to update their contact information using a web address update form. Conducting panel maintenance activities allows us to maintain up-to-date contact information which increases the likelihood of locating sample members, reduces the time elapsed between contacts, and maximizing response rates for the subsequent follow-up. Approval to conduct the panel maintenance activities on the BPS:12/17 full-scale sample was obtained through the BPS:12/17 Pilot Test submission (OMB #1850-0803 v.150). The initial contact letter and email, and the web address update form, may be found in Appendix E.

b. Prompting Cases with Mail, Email, and SMS Contacts

Past experience on recent postsecondary studies, including the BPS:12/17 pilot test, has shown that maintaining frequent contact with sample members through mailings, emails, and text messages sent at regular intervals improves response rates. The reminders that will be regularly sent throughout the study include:

- Initial contact letter and email sent to introduce the study and request updated contact information for the sample member.
- Data collection announcement letter and emails sent to announce the start of the data collection period and encourage sample members to complete the web interview.
- Reminder letters and postcards sent to nonresponding sample members throughout data collection and customized to address to concerns of nonresponding sample members.
- Brief reminder emails to remind nonresponding sample members to complete the interview.
- Brief text (SMS) message reminders sent to those sample members who have provided consent to receive text messages.
- Brief letters and emails sent after the completion of the interview that express gratitude for participation, tell the respondent when to anticipate the incentive, and instruct the respondent on how to contact us with any concerns about the incentive.

Examples of the planned contacting materials are also included in Appendix E.

c. Telephone Interviewing and Help Desk Support

Telephone interviewer training. Well-trained interviewers play a critical role in gaining sample member cooperation. The BPS:12/17 training program will include an iLearning module to provide staff with a study introduction prior to in-person training; 12 hours of in-person study training with mock interviews, hands-on practice, and other exercises; and ongoing training throughout data collection. The BPS:12/17

Telephone Interviewer manual will cover the background and purpose of BPS:12/17, instructions for providing Help Desk support, and procedures for administering the telephone interview. At the conclusion of training, all interviewers must meet certification requirements by successfully completing a full-length certification interview and an oral test of the study's Frequently Asked Questions, and by demonstrating proper pronunciation of study terms. Interviewers will be expected to knowledgably and extemporaneously respond to sample members' questions about the study, thereby increasing their chances of gaining cooperation.

Interviewing and Prompting. Interviews will be conducted using a single, web-based survey instrument for all modes of data collection—self-administered (via computer or mobile device) or interviewer-administered (via telephone interviewers). Data collection for the calibration study will begin in February 2017, and calibration sample members will be encouraged to complete the survey on the web during the initial 4-week early completion period, followed by outbound calling for sample members who did not respond during the early completion period.

Telephone interviewing and prompting will be managed using the CATI Case Management System (CMS). The CATI-CMS is equipped to provide:

- Complete records of locating information and histories of locating and contacting efforts for each case;
- Reporting capabilities, including default reports on the aggregate status of cases and custom reports defined as needed;
- An automated scheduling module, which provides highly efficient case assignment and delivery functions, reduces supervisory and clerical time, and improves execution on the part of interviewers and supervisors by automatically monitoring appointments and callbacks. The scheduler delivers cases to telephone interviewers and incorporates the following efficiency features:
 - automatic delivery of appointment and call-back cases at specified times;
 - sorting of non-appointment cases according to parameters and priorities set by project staff;
 - complete records of calls and tracking of all previous outcomes; and
 - tracking of problem cases for supervisor action or supervisor review.

For most cases, outbound prompting will begin 4 weeks after the start of data collection. As with NPSAS:12 and BPS:12/14, RTI plans to initiate telephone prompting efforts earlier for some challenging cases to maximize response rates. These may include NPSAS:12 and BPS:12/14 nonrespondents and sample members from institutional sectors with historically low participation rates. For these cases, outbound prompting may begin as early as 2 weeks after the start of data collection.

Refusal Aversion and Conversion. Recognizing and avoiding refusals is important to maximizing the response rate for BPS:12/17. All interviewers will be trained in techniques that are designed to gain cooperation and avoid refusals whenever possible. BPS training program stresses the importance of learning the most frequently asked questions. When a sample member has concerns about participation, interviewers are expected to respond to these questions knowledgably and confidently.

Supervisors will monitor interviewers intensively during the early weeks of data collection and provide retraining as necessary. In addition, supervisors will review daily interviewer production reports to identify and retrain any interviewers with unacceptable numbers of refusals or other problems. After encountering a refusal, comments are entered into the CMS record that include all pertinent data regarding the refusal situation, including any unusual circumstances and any reasons given by the sample member for refusing. Supervisors review these comments to determine what action to take with each refusal. No refusal or partial interview will be coded as final without supervisory review and approval.

If follow-up to an initial refusal is not appropriate (e.g., there are extenuating circumstances such as illness or the sample member firmly requested no further contact), the case will be coded as final and no

additional contact will be made. If the case appears to be a “soft” refusal (i.e., an initial refusal that warrants conversion efforts – such as expressions of too little time or lack of interest or a telephone hang-up without comment), follow-up will be assigned to a member of a special refusal conversion team of interviewers skilled at converting refusals. Refusal conversion efforts will be delayed until at least 1 week 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. Project staff sometimes receive refusals via email or calls to the project toll-free line. These refusals are included in the CATI record of events and coded as final when appropriate.

d. Quality Control

Interviewer monitoring will be conducted using the Quality Evaluation System (QUEST) as a quality control measure throughout the data collection. QUEST is a system developed by a team of RTI researchers, methodologists, and operations staff focused on developing standardized monitoring protocols, performance measures, evaluation criteria, reports, and appropriate system security controls. It is a comprehensive performance quality monitoring system that includes standard systems and procedures for all phases of the interview process, including (a) obtaining respondent consent for recording, interviewing respondents who refuse consent for recording, and monitoring refusals at the interviewer level; (b) sampling of completed interviews by interviewer and evaluating interviewer performance; (c) maintaining an online database of interviewer performance data; and (d) addressing potential problems through supplemental training.

As in previous studies, calls will be reviewed by call center supervisors for key elements such as professionalism and presentation; case management and refusal conversion; and reading, probing, and keying skills. Feedback will be provided to interviewers and patterns of poor performance will be carefully documented, and if necessary, addressed with additional training. Sample members will be notified that the interview may be monitored by supervisory staff.

Regular Quality Circle meetings will provide another opportunity to ensure quality and consistency in interview administration. These meetings give interviewers a forum to ask questions about best practices and share their experiences with other interviewers. They provide opportunities for brainstorming strategies for avoiding or converting refusals and optimizing interview quality, which ensures a positive experience for sample members.

e. Administrative Record Collections

The following sections describes the approach that we will employ in order to efficiently collect transcripts and student records from institutions. The general approach is described first, followed by procedures specific to each collection.

The success of the BPS:12 PETS and Student Records is dependent on the active participation of sampled institutions. The cooperation of an institution’s coordinator is essential, and helps to encourage the timely completion of the data collection. Telephone contact between the project team and institution coordinators provides an opportunity to emphasize the importance of the study and to address any concerns about participation.

BPS:12 procedures for working with institutions will be developed from those used successfully in other studies with collections of transcripts and records, such as the ELS:2002, NPSAS:12, and NPSAS:16. We will use an institution control system (ICS) currently used for NPSAS:16 and similar to the system used for ELS:2002 to maintain relevant information about the institutions attended by each BPS:12 cohort member. Institution contact information obtained from the Integrated Postsecondary Education Data System (IPEDS) will be loaded into the ICS and confirmed during a call to the institutional research (IR) office or general operator. The initial request mailing will be sent to IR director, or in the absence of an IR

director, a chief administrator. In the initial call to the IR director, he/she will be asked to designate a primary coordinator for the data collection.

In past studies, the specific endorsement of relevant associations and organizations has been extremely useful in persuading institutions to cooperate. Endorsements from 16 professional associations were secured for ELS:2002. Appropriate associations to request endorsement for BPS:12 PETS and Student Records collections will be contacted as well; a list of potential endorsing associations and organizations is provided in appendix F.

Another successful strategy has been to solicit support at a system-wide level rather than contacting each institution within the system. A timely contact, together with enhanced institution contact information verification procedures, is likely to reduce the number of re-mail requests, and minimize delay caused by misrouted requests.

In addition, NCES plans to conduct BPS:12 PETS and Student Records at the same time as, and in collaboration with, institution data collections for HSL:09. The combined collections will reduce burden on institutions by minimizing the number of NCES requests for data and by using the same tools and procedures for institutions' reporting. Requests for administrative data will include students from both studies and guide institution representatives to the NCES Postsecondary Data Portal (PDP), a one-stop website where institutions can provide data for students at the same time, regardless of the NCES postsecondary sample study, using the same submission layout and functionality across the studies. The PDP has been designed to be user-friendly and offers several modes for providing data so that each institution can choose their preferred way. The PDP is located at <https://surveys.nces.ed.gov/Portal/>.

The combined data collection approach will also reduce institution contacting costs (by reducing printing and shipping of materials) and labor costs (by using overlapping project staff and the same set of institution contactors). The current plan is to conduct BPS:12 Student Records and PETS pilot test at the same time as HSL:09 transcripts and student records collection (approved in March 2016; OMB# 1850-0852) and, if feasible, to conduct the BPS:12 PETS collection at the same time as the NPSAS:18 student records collection (for which NCES will submit OMB clearance request in mid-2017).

At the start of data collection, each institution will be sent a non-study-specific packet containing a letter and other information about the PDP. The second packet sent will be a data request that specifies the type of data being requested and the studies for which the request is being made. The materials are provided in appendix F and include the following:

- An introductory letter on NCES letterhead;
- A letter from RTI providing information on the data collections;
- Letter(s) of endorsement from supporting organizations/agencies (e.g., the American Association of Collegiate Registrars and Admission Officers);
- A list of other endorsing agencies;
- Directions regarding how to log on to PDP, how to access the list of students for whom information is being requested, and how to request reimbursement of incurred expenses (e.g., transcript processing fees); and
- Descriptions of and instructions for the various methods of providing transcripts and student records data.

In the same materials, institutions will be asked to participate in both the PETS and Student Records collections, if both apply to the institution. If separate institution staff are identified to provide different types of data, the contact materials will be provided to those staff members as needed. Follow-up contacts will occur after the initial mailing to ensure receipt of the package and to answer any questions about the study, as applicable.

BPS:12 PETS Collection. A complete transcript from the institution will be requested as well as the complete transcripts from transfer institutions that the students attended, as applicable. A Transcript Control System (TCS) will track receipt of institution materials and student transcripts for BPS:12 PETS and HSLs:09. The TCS will track the status of each catalog and transcript request, from initial mailout of the request through follow-ups and final receipt.

As described in section A.10, active student consent for the release of transcripts will not be required. In compliance with FERPA, a notation will be made in the student record that the transcript has been collected for use for statistical purposes only in the BPS:12 and/or HSLs:09 longitudinal study(ies).

The following methods will be used for institutions to deliver the requested transcripts:

- *PDP.* Because of the risks associated with transmitting confidential data on the internet, the latest technology systems will be incorporated into the web application to ensure strict adherence to NCES confidentiality guidelines. The web server will include a Secure-Sockets Layer (SSL) Certificate, and will be configured to force encrypted data transmission over the Internet. All data entry modules on this site will be password protected, requiring the user to log in to the site before accessing confidential data. To access restricted pages containing confidential information, the user will be required to log in by entering an assigned ID number and password. Through the PDP, the primary coordinators at the institution will be able to use a 'Manage Users' link, available only to them, to add and delete users, as well as reset passwords and assign roles. Each user will have a unique username and will be assigned to one email address. Upon account creation, the new user will be sent a temporary password by the PDP. Upon logging in for the first time, the new user will be required to create a new password. The system automatically will log out the user after 20 minutes of inactivity. Files uploaded to the secure website will be stored in a secure project folder that is only accessible and visible to authorized project staff.
- *File Transfer Protocol.* FTPS (also called FTP-SSL) uses the FTP protocol on top of SSL or Transcript Layer Security (TLS). When using FTPS, the control session is always encrypted. The data session can optionally be encrypted if the file has not already been encrypted. Files transmitted via FTPS will be placed in a secure project folder that is only accessible and visible to authorized staff members.
- *Encrypted attachments via email.* RTI will provide guidelines on encryption and on creating strong passwords. Encrypted electronic files sent via email to a secure email folder will only be accessible to a limited set of authorized staff members on the project team. These files will then be copied to a project folder that is only accessible and visible to these same staff members.
- *eSCRIP-SAFE™.* This method involves the institution sending data via a customized print driver which connects the student information system to the eSCRIP-SAFE™ server by secure internet connection. RTI, as the designated recipient, can then download the data after entering a password. The files are deleted from the server 24 hours after being accessed. The transmission between sending institutions and the eSCRIP-SAFE™ server is protected by SSL connections using 128-bit key ciphers. Remote access to the eSCRIP-SAFE™ server via the web interface is likewise protected via 128-bit SSL. Downloaded files will be moved to a secure project folder that is only accessible and visible to authorized staff members.
- *Secure electronic fax (e-fax).* We expect that a few institutions will ask to provide hardcopy transcripts. If more secure options are not possible, faxed transcripts will be accepted. Although fax equipment and software does facilitate rapid transmission of information, this same equipment and software opens up the possibility that information could be misdirected or intercepted by unauthorized individuals. To safeguard against this, we will only allow for transcripts to be faxed to an electronic fax machine and only in the absence of other options. To ensure transmission to the intended destination, a test fax with non-sensitive data will be required to reduce errors in

transmission from misdialing. Institutions will be given a fax cover page that includes a confidentiality statement to use when transmitting individually identifiable information. Transcripts received via e-fax are temporarily stored on the e-fax server, which is housed in a secured data center at RTI. These files will be copied to a project folder that is only accessible and visible to authorized staff members.

- *Federal Express*. When institutions ask to provide hardcopy transcripts, they will be encouraged to use one of the secure electronic methods of transmission or fax. If that is not possible, transcripts sent via Federal Express will be accepted. Before sending, institution staff will be instructed to redact any personally identifiable information from the transcript including student name, address, date of birth, and Social Security Number (if present). Paper transcripts will be scanned and stored as electronic files. These files will be stored in a project folder that is only accessible and visible to project staff members. The original paper transcripts will be shredded.
- *National Student Clearinghouse*. Approximately 200 institutions are currently registered to send and receive academic transcripts in standardized electronic formats via the SPEEDE dedicated server. Additional institutions are in the test phase. The SPEEDE server supports the following methods of securely transmitting transcripts: email as MIME attachment using Pretty Good Privacy (PGP) encryption; regular FTP using PGP encryption; Secure FTP (SFTP over SSH) and straight SFTP; and FTPS (FTP over SSL/TLS). Files collected via this dedicated server will be copied to a secure project folder that is only accessible and visible to authorized staff members. The same access restrictions and storage protocol will be followed for these files as described above for files uploaded to the PDP.

As part of quality control procedures, the importance of collecting complete transcript information for all sampled students will be emphasized to registrars. Transcripts will be reviewed for completeness. Institutional contactors will contact the institutions to prompt for missing data and to resolve any problems or inconsistencies.

Transcripts received in hardcopy form will be subject to a quick review prior to recording their receipt. Receipt control clerks will check transcripts for completeness and review transmittal documents to ensure that transcripts have been returned for each of the specified sample members. The disposition code for transcripts received will be entered into the TCS. Course catalogs will also be reviewed and their disposition status updated in the system in cases where this information is necessary and not available through CollegeSource Online. Hardcopy course catalogs will be sorted and stored in a secure facility at RTI, organized by institution. The procedures for electronic transcripts will be similar to those for hardcopy documents—receipt control personnel, assisted by programming staff, will verify that the transcript was received for the given requested sample member, record the information in the receipt control system, and check to make sure that a readable, complete electronic transcript has been received.

The initial transcript check-in procedure is designed to efficiently log the receipt of materials into the TCS as they are received each day. The presence of an electronic catalog (obtained from CollegeSource Online) will be confirmed during the verification process for each institution and noted in the TCS. The remaining catalogs will be requested from the institutions directly and will be logged in the TCS as they are received. Transcripts and supplementary materials received from institutions (including course catalogs) will be inventoried, assigned unique identifiers based on the IPEDS ID, reviewed for problems, and logged into the TCS.

Project staff will use daily monitoring reports to review the transcript problems and to identify approaches to solving the problems. The web-based collection will allow timely quality control, as RTI staff will be able to monitor data quality for participating institutions closely and on a regular basis. When institutions call for technical or substantive support, the institution's data will be queried in order to communicate with the institution more effectively regarding problems. Transcript data will be destroyed or shredded after the transcripts are keyed, coded, and quality checked.

A comprehensive supervision and quality control plan will be implemented during transcript keying and coding. At least one supervisor will be onsite at all times to manage the effort and simultaneously perform QC checks and problem resolution. Verifications of transcript data keying and coding at the student level will be performed. Any errors will be recorded and corrected as needed.

BPS:12 Student Records Collection. Institution coordinators will receive a guide that provides instructions for accessing and using the PDP (as described above). In conjunction with the transcript collection, RTI institution contacting staff will notify institutions that the student records data collection has begun and will follow up (by telephone, mail, or email) as needed. Project staff will be available by telephone and email to provide assistance when institution staff have questions or encounter problems.

The following options will be offered to institutions for collecting student records:

- *Web-based data entry interface.* The web-based data entry interface allows the coordinator to enter data by student, by year.
- *Excel workbook.* An Excel workbook will be created for each institution and will be preloaded with each sampled student's ID, name, date of birth, and last four digits of SSN (if available). To facilitate simultaneous data entry by different offices within the institution, the workbook contains a separate worksheet for each of the following topic areas: Student Information, Financial Aid, Enrollment, and Budget. The user will download the Excel worksheet from the PDP, enter the data, and then upload the data. Validation checks will occur both within Excel as data are entered and when the data are uploaded. Data will be imported into the web application so that institution staff can check their submission for quality control purposes.
- *CSV (comma separated values) file.* Institutions with the means to export data from their internal database systems to a flat file may use this method of supplying student records. Institutions that select this method will be provided with detailed import specifications, and all data uploading will occur through the PDP. Like the Excel workbook option, data will be imported into the web application such that institution staff can check their submission before finalizing.

4. Tests of Procedures and Methods

The design of the BPS:12/17 full-scale data collection—in particular, the use of responsive design principles to reduce bias associated with nonresponse—expands on data collection experiments designed for several preceding NCES studies and, particularly, on the responsive design methods employed in BPS:12/14. Section B.4.a below provides an overview of the responsive design methods employed for BPS:12/14, section B.4.b provides a description of the proposed methods for BPS:12/17, and section B.4.c describes the tests that will be conducted through the BPS:12 PETS pilot study.

a. *BPS:12/14 Full Scale*³

The BPS:12/14 full-scale data collection combined two experiments in a responsive design (Groves and Heeringa 2006) in order to examine the degree to which targeted interventions could affect response rates and reduce nonresponse bias. Key features included a calibration sample for identifying optimal monetary incentives and other interventions, the development of an importance measure for use in identifying nonrespondents for some incentive offers, and the use of a six-phase data collection period.

³ This section addresses the following BPS terms of clearance: (1) From OMB# 1850-0631 v.8: “OMB approves this collection under the following terms: At the conclusion of each of the two monetary incentive calibration activities, NCES will meet with OMB to discuss the results and to determine the incentive amounts for the remaining portion of the study population. Further, NCES will provide an analytical report back to OMB of the success, challenges, lessons learned and promise of its approach to addressing non-response and bias via the approach proposed here. The incentive levels approved in this collection do not provide precedent for NCES or any other Federal agency. They are approved in this specific case only, primarily to permit the proposed methodological experiments.”; and (2) From OMB# 1850-0631 v.9: “Terms of the previous clearance remain in effect. NCES will provide an analytical report back to OMB of the success, challenges, lessons learned and promise of its approach to addressing non-response and bias via the approach proposed here. The incentive levels approved in this collection do not provide precedent for NCES or any other Federal agency. They are approved in this specific case only, primarily to permit the proposed methodological experiments.”

Approximately 10 percent of the 37,170 BPS:12/14 sample members were randomly selected to form the *calibration sample*, with the remainder forming the *main sample*, although readers should note that respondents from the calibration and main sample were combined at the end of data collection. Both samples were subject to the same data collection activities, although the calibration sample was fielded seven weeks before the main sample.

First Experiment: Determine Baseline Incentive. The first experiment with the calibration sample, which began with a web-only survey at the start of data collection (**Phase 1**), evaluated the baseline incentive offer. In order to assess whether or not baseline incentive offers should vary by likelihood of response, an *a priori* predicted probability of response was constructed for each calibration sample member. Sample members were then ordered into five groups using response probability quintiles and randomly assigned to one of eleven baseline incentive amounts ranging from \$0 to \$50 in five dollar increments. Additional information on how the *a priori* predicted probabilities of response were constructed is provided below.

For the three groups with the highest predicted probabilities of response, response rates for a given baseline incentive offer (\$0 to \$25) were statistically higher than response rates for the next lowest incentive amount up to \$30. In addition, response rates for incentives of \$35 or higher were not statistically higher than response rates at \$30. For the two groups with the lowest predicted probabilities of response, the response rate at \$45 was found to be statistically higher than the response rate at \$0, but the finding was based on a small number of cases. Given the results across groups, a baseline incentive amount of \$30 was set for use with the main sample. Both calibration and main sample nonrespondents at the end of Phase 1 were moved to **Phase 2** with outbound calling; no changes were made to the incentive level assigned at the start of data collection.

Second Experiment: Determine Monetary Incentive Increase. Phase 3, a second experiment implemented with the calibration sample, after the first 28 days of Phase 2 data collection, determined the additional incentive amount to offer the remaining nonrespondents with the highest “value” to the data collection, as measured by an “importance score” (see below). During Phase 3, 500 calibration sample nonrespondents with the highest importance scores were randomly assigned to one of three groups to receive an incentive boost of \$0, \$25, or \$45 in addition to the initial offer.

Across all initial incentive offers, those who had high importance scores but were in the \$0 incentive boost group had a response rate of 14 percent, compared to 21 percent among those who received the \$25 incentive boost, and 35 percent among those who received the \$45 incentive boost. While the response rate for the \$25 group was not statistically higher than the response rate for the \$0 incentive group, the response rate for the \$45 group was statistically higher than the response rates of both the \$25 and the \$0 groups. Consequently, \$45 was used as the additional incentive increase for the main sample.

Importance Measure. Phases 1 and 3 of the BPS:12/14 data collection relied on two models developed specifically for this collection. The first, an *a priori* response propensity model, was used to predict the probability of response for each BPS:12/14 sample member prior to the start of data collection (and assignment to the initial incentive groups). Because the BPS:12/14 sample members were part of the NPSAS:12 sample, predictor variables for model development included sampling frame variables and NPSAS:12 variables including, but not limited to, the following:

- responded during early completion period,
- interview mode (web/telephone),
- ever refused,
- call count, and
- tracing/locating status (located/required intensive tracing).

The second model, a bias-likelihood model, was developed to identify those nonrespondents, at a given point during data collection, who were most likely to contribute to nonresponse bias. At the beginning of

Phase 3, described above, and of the next two phases – local exchange calling (Phase 4) and abbreviated interview for mobile access (Phase 5) – a logistic regression model was used to estimate, not predict, the probability of response for each nonrespondent at that point. The estimated probabilities highlight individuals who have underrepresented characteristics among the respondents at the specific point in time. Variables used in the bias-likelihood model were derived from base-year (NPSAS:12) survey responses, school characteristics, and sampling frame information. It is important to note that paradata, such as information on response status in NPSAS:12, particularly those variables that are highly predictive of response but quite unrelated to the survey variables of interest, were excluded from the bias-likelihood model. Candidate variables for the model included:

- highest degree expected,
- parents' level of education,
- age,
- gender,
- number of dependent children,
- income percentile,
- hours worked per week while enrolled,
- school sector,
- undergraduate degree program,
- expected wage, and
- high school graduation year.

Because the variables used in the bias-likelihood model were selected due to their potential ability to act as proxies for survey outcomes, which are unobservable for nonrespondents, the predicted probabilities from the bias-likelihood model were used to identify nonrespondents in the most underrepresented groups, as defined by the variables used in the model. Small predicted probabilities correspond to nonrespondents in the most underrepresented groups, i.e. most likely to contribute to bias, while large predicted probabilities identify groups that are, relatively, well-represented among respondents.

The importance score was defined for nonrespondents as the product of a sample member's *a priori* predicted probability of response and one minus the sample member's predicted bias-likelihood probability. Nonrespondents with the highest calculated importance score at the beginning of Phases 3, 4, and 5, were considered to be most likely to contribute to nonresponse bias and, therefore, were offered the higher monetary incentive increase (Phase 3), were sent to field and local exchange calling (Phase 4), and were offered an abbreviated interview (Phase 5). An overview of the calibration and main sample data collection activities is provided in table 5.

Table 5. Summary of start dates and activities for each phase of the BPS:12/14 data collection, by sample

Phase	Start date		Activity	
	Calibration subsample	Main subsample	Calibration subsample	Main subsample
1	2/18/2014	4/8/2014	Begin web collection; Randomize calibration sample to different baseline incentives (experiment #1)	Begin web collection; baseline incentives determined by results of first calibration experiment
2	3/18/2014	5/6/2014	Begin CATI collection	Begin CATI collection
3	4/8/2014	5/27/2014	Randomize calibration sample nonrespondents to different monetary incentive increases (experiment #2)	Construct importance score and offer incentive increase to select nonrespondents; incentive increase determined by results of second calibration experiment
4	5/6/2014	6/24/2014	Construct importance score and identify select nonrespondents for Field/local exchange calling for targeted cases	Construct importance score and identify select nonrespondents for Field/local exchange calling for targeted cases
5	7/15/2014	9/2/2014	Construct importance score and identify select nonrespondents for abbreviated interview with mobile access	Construct importance score and identify select nonrespondents for abbreviated interview with mobile access
6	8/12/2014	9/30/2014	Abbreviated interview for all remaining nonrespondents	Abbreviated interview for all remaining nonrespondents

CATI = computer-assisted telephone interviewing

Impact on Nonresponse Bias. As all BPS:12/14 sample members were submitted to the same data collection procedures, there is no exact method to assess the degree to which the responsive design reduced nonresponse bias relative to another data collection design that did not incorporate responsive design elements. However, a post-hoc analysis was implemented to compare estimates of nonresponse bias to determine the impact of the responsive design. Nonresponse bias estimates were first created using all respondents and then created again by reclassifying targeted respondents as nonrespondents. This allows examination of the potential bias contributed by the subset of individuals who were targeted by responsive design methods although this is not a perfect design as some of these individuals would have responded without interventions. The following variables were used to conduct the nonresponse bias analysis:⁴

- Region (categorical);
- Age as of NPSAS:12 (categorical);
- CPS match as of NPSAS:12 (yes/no);
- Federal aid receipt (yes/no);
- Pell Grant receipt (yes/no);
- Pell Grant amount (categorical);
- Stafford Loan receipt (yes/no);
- Stafford Loan amount (categorical);
- Institutional aid receipt (yes/no);
- State aid receipt (yes/no);
- Major (categorical);
- Institution enrollment from IPEDS file (categorical);
- Any grant aid receipt (categorical); and
- Graduation rate (categorical).

⁴ For the continuous variables, except for age, categories were formed based on quartiles.

For each variable listed above, nonresponse bias was estimated by comparing estimates from base-weighted respondents with those of the full sample to determine if the differences were statistically significant at the 5 percent level. Multilevel categorical terms were examined using indicator terms for each level of the main term. The relative bias estimates associated with these nonresponse bias analyses are summarized in Table 6.

The mean and median percent relative bias are almost universally lowest across all sectors when all respondents are utilized in the bias assessment. The overall percentage of characteristics with significant bias is lowest when all respondents are utilized but the percentage of characteristics with significant bias is lowest in seven of the ten sectors when responsive design respondents are excluded. However, the percentage of characteristics with significant bias is affected by sample sizes and as there are approximately 5,200 respondents who were ever selected under the responsive design, the power to detect a bias that is statistically different from zero is higher when using all respondents versus a smaller subset of those respondents in a nonresponse bias assessment. Consequently, the mean and median percent relative bias are better gauges of how the addition of selected responsive design respondents impacts nonresponse bias.

Given that some of the 5,200 selected respondents would have responded even if they had never been subject to responsive design, it is impossible to attribute the observed bias reduction solely to the application of responsive design methods. However, observed reduction of bias is generally quite large and suggests that responsive design methods may be helpful in reducing nonresponse bias.

Table 6. Summary of responsive design impact on nonresponse bias, by institutional sector: 2014

Nonresponse bias statistics¹	Overall	Public less-than- 2-year	Public 2-year	Public 4-year non- doctorate- granting	Public 4-year doctorate- granting	Private nonprofit less-than- 4-year	Private nonprofit 4-year non- doctorate- granting	Private nonprofit 4-year doctorate- granting	Private for-profit less-than- 2-year	Private for-profit 2-year	Private for-profit 4-year
All respondents											
Mean percent relative bias across characteristics	10.7	7.8	6.4	8.9	4.2	12.9	7.1	4.7	13.0	9.6	7.2
Median percent relative bias across characteristics	6.3	5.1	3.8	4.6	2.7	8.8	4.0	3.8	7.5	5.4	5.5
Percentage of characteristics with significant bias	62.1	36.4	29.0	43.2	33.3	7.4	32.6	29.3	25.0	8.3	11.7
Respondents excluding those selected for responsive design											
Mean percent relative bias across characteristics	22.9	12.5	12.8	13.6	4.3	29.0	16.5	4.7	12.1	18.2	11.6
Median percent relative bias across characteristics	12.4	9.7	5.0	7.8	3.0	16.5	4.2	3.6	4.7	9.0	8.1
Percentage of characteristics with significant bias	73.6	16.7	26.9	55.6	37.5	32.4	31.8	30.2	17.1	5.7	9.7

¹ Relative bias and significance calculated on respondents vs. full sample.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2012/14 Beginning Postsecondary Students Longitudinal Study (BPS:12/14).

b. *BPS:12/17 Full Scale*

The responsive design methods proposed for BPS:12/17 expand and improve upon the BPS:12/14 methods in three key aspects:

- 1) Refined targeting of nonresponding sample members so that, instead of attempting to reduce unit nonresponse bias for national estimates only, as in BPS:12/14, the impact of unit nonresponse on the bias is reduced for estimates *within institutional sector*.
- 2) Addition of a special data collection protocol for a hard-to-convert group: NPSAS:12 study member double-interview nonrespondents.
- 3) Inclusion of a randomized evaluation designed to permit estimating the difference between unit nonresponse bias arising from application of the proposed responsive design methods and unit nonresponse bias arising from not applying the responsive design methods.

As noted previously, the responsive design approach for the BPS:12/14 full scale included (1) use of an incentive calibration study sample to identify optimal monetary incentives, (2) development of an importance measure for identifying nonrespondents for specific interventions, and (3) implementation of a multi-phase data collection period. Analysis of the BPS:12/14 case targeting indicated that institution sector dominated the construction of the importance scores; meaning that nonrespondents were primarily selected by identifying nonrespondents in the sectors with the lowest response rates. For the BPS:12/17 full scale we are building upon the BPS:12/14 full scale responsive design but, rather than selecting nonrespondents using the same approach as in BPS:12/14, we propose targeting nonrespondents within:

- Institution Sector – we will model and target cases within sector groups in an effort to equalize response rates across sectors.
- NPSAS:12 study member double interview nonrespondents – we will use a calibration sample to evaluate two special data collection protocols for this hard-to-convert group, including a special baseline protocol determined by a calibration sample and an accelerated timeline.

We have designed an evaluation of the responsive design so that we can test the impact of the targeted interventions to reduce nonresponse bias versus not targeting for interventions. For the evaluation, we will select a random subset of all sample members to be pulled aside as a *control* sample that will not be eligible for intervention targeting. The remaining sample member cases will be referred to as the *treatment* sample and the targeting methods will be applied to that group.

In the following sections, we will describe the proposed importance measure, sector grouping, and intervention targeting, then describe the approach for the pre-paid and double nonrespondent calibration experiments, and outline how these will be implemented and evaluated in the BPS:12/17 full scale data collection.

The importance measure. In order to reduce nonresponse bias in survey variables by directing effort and resources during data collection, and to minimize the cost associated with achieving this goal, three related conditions have to be met: (1) the targeted cases must be drawn from groups that are under-represented on key survey variable values among those who already responded, (2) their likelihood of participation should not be excessively low or high (i.e., targeted cases who do not respond cannot decrease bias; targeting only high propensity cases can potentially increase the bias of estimates), and (3) targeted cases should be numerous enough to impact survey estimates within domains of interest. While targeting cases based on response propensities may reduce nonresponse bias, bias may be unaffected if the targeted cases are extremely difficult to convert and do not respond to the intervention as desired.

One approach to meeting these conditions is to target cases based on two dimensions: the likelihood of a case to contribute to nonresponse bias if not interviewed, and the likelihood that the case could be converted to a respondent. These dimensions form an importance score, such that:

$$I_{ij} \propto U_{ij} \cdot P(R)_i$$

Where I is the calculated importance score, U is a measure of under-representativeness on key variables that reflects their likelihood to induce bias if not converted, and $P(R)$ is the predicted final response propensity, across i sample members and j data collection phases with responsive design interventions.

The importance score will be determined by the combination of two models: a response propensity model and a bias-likelihood model. Like BPS:12/14, the response propensity component of the importance score is being calculated in advance of the start of data collection. The representativeness of key variables, however, can only be determined during specific phases of the BPS:12/17 data collection, with terms tailored to BPS:12/17. The importance score calculation needs to balance two distinct scenarios: (1) low propensity cases that will likely never respond, irrespective of their underrepresentation, and (2) high propensity cases that, because they are not underrepresented in the data, are unlikely to reduce bias. Once in production, NCES will provide more information about the distribution of both propensity and representation from the BPS:12/17 calibration study, which will allow us to explore linear and nonlinear functions that optimize the potential for nonresponse bias and available incentive resources. We will share the findings with OMB at that time.

Bias-likelihood (U) model. A desirable model to identify cases to be targeted for intervention would use covariates (Z) that are strongly related to the survey variables of interest (Y), to identify sample members who are under-represented (using a response indicator, R) with regard to these covariates. We then have the following relationships, using a single Z and Y for illustration:



Nonresponse bias arises when there is a relationship between R and Y . Just as in adjustment for nonresponse bias (see Little and Vartivarian, 2005), a Z -variable cannot be effective in nonresponse bias reduction if $\text{corr}(Z, Y)$ is weak or nonexistent, even if $\text{corr}(Z, R)$ is substantial. That is, selection of Z -variables based only on their correlation with R may not help to identify cases that contribute to nonresponse bias. The goal is to identify sample cases that have Y -variable values that are associated with lower response rates, as this is one of the most direct ways to reduce nonresponse bias in an estimate of a mean.

The key Z -variable selection criterion should then be association with Y . Good candidate Z -variables would be the Y -variables or their proxies measured in a prior wave and any correlates of change in estimates over time. A second set of useful Z -variables would be those used in weighting and those used to define subdomains for analysis – such as demographic variables. This should help to reduce the variance inflation due to weighting and nonresponse bias in comparisons across groups. Key, however, is the exclusion of variables that are highly predictive of R , but quite unrelated to Y . These variables, such as the number of prior contact attempts and prior refusal, can dominate in a model predicting the likelihood of participation and mask the relationship of Z variables that are associated with Y .

Prior to the start of later phases of data, when the treatment interventions will be introduced, we will conduct multiple logistic regressions in order to predict the survey outcome (R) through the current phase of collection using only substantive and demographic variables and their correlates from NPSAS:12 and the sampling frame (Z), and select two-way interactions. For each sector grouping (see table 8 below), a single model will be fit. The goal of this model is not to maximize the ability to predict survey response (\hat{p}), but to obtain a predicted likelihood of a completed interview reducing nonresponse bias if successfully interviewed. Because of this key difference, we use $(1 - \hat{p})$ to calculate a case-level prediction representing bias-likelihood, rather than response propensity.

Variables to be used in the bias-likelihood model will come from base-year survey responses, institution characteristics, and sampling frame information⁵ (see table 7). It is important to note that paradata, particularly those variables that are highly predictive of response, but quite unrelated to the survey variables of interest, will be excluded from the bias-likelihood model.

Table 7. Candidate variables for the bias likelihood model

Variables	
Race	Attendance intensity
Gender	Highest level of education ever expected
Age	Dependent children and marital status
Sector*	Federal Pell grant amount
Match to Central Processing System	Direct subsidized and unsubsidized loans
Match to Pell grant system	Total federal aid
Total income	Institutional aid total
Parent's highest education level	Degree program

* Variable to be included in bias likelihood model for targeting sample members from public 4-year and private nonprofit institutions (sector group B in table 8).

Response propensity (P(R)) model. Prior to the start of BPS:12/17 data collection, a response propensity model is being developed to predict likelihood to respond to BPS:12/17 based on BPS:12/14 data and response behavior. NCES will share the model with OMB when finalized and prior to implementation. The model will use variables from the base NPSAS:12 study as well as BPS:12/14 full scale that have been shown to predict survey response, including, but not limited to:

- responded during early completion period,
- response history,
- interview mode (web/telephone),
- ever refused,
- incentive amount offered,
- age,
- gender,
- citizenship,
- institution sector,
- call count, and
- tracing/locating status (located/required intensive tracing).

We will use BPS:12/14 full scale data to create this response propensity model as that study was similar in design and population to the current BPS:12/17 full scale study (note that BPS:12/17 did not have a field test that could be leveraged, and the pilot study was too limited in size and dissimilar in approach and population to be useful for this purpose).

Targeted interventions. In the BPS:12/14 responsive design approach, institution sector was the largest factor in determining current response status. For BPS:12/17 full scale, individuals will be targeted within groupings of institution sectors in an effort to equalize response rates across the sector groups. Designed to reduce the final unequal weighting effect, targeting within the groups will allow us to fit a different propensity or bias likelihood model for each group while equalizing response rates across groups.

Targeting within sector groups is designed to reduce nonresponse bias within specific sectors rather than across the aggregate target population. The five sector groupings (Table 8) were constructed by first identifying sectors with historically low response rates, as observed in BPS:12/14 and NPSAS:12, and, second, assigning the sectors with the lowest participation to their own groups. The remaining sectors were then combined into groups consisting of multiple sectors. The private for profit sectors (groups C, D, and E) were identified to have low response rates. Public less-than-2-year and public 2-year institutions (group A) were combined as they were similar, and because the public less-than-2-year sector was too small to act as a

⁵ Key variables will use imputed data to account for nonresponse in the base year data.

distinct group. Public 4-year and private nonprofit institutions (sector group B) remained combined as they have not historically exhibited low response rates (nonetheless, cases within this sector group are still eligible for targeting; the targeting model for sector group B will include sector as a term to account for differences between the sectors).

Table 8. Targeted sector groups

Sector Group	Sectors	Sample Count
A	1: Public less-than-2-year	172
	2: Public 2-year	10,030
B	3: Public 4-year non-doctorate-granting	1,838
	4: Public 4-year doctorate-granting	3,545
	5: Private nonprofit less than 4-year	397
	6: Private nonprofit 4-year nondoctorate	2,219
	7: Private nonprofit 4-year doctorate-granting	2,717
C	8: Private for profit less-than-2-year	1,450
D	9: Private for profit 2-year	2,989
E	10: Private for profit 4-year	8,401

All NPSAS:12 study members who responded to the NPSAS:12 or BPS:12/14 student interviews (hereafter called previous respondents) will be initially offered a \$30 incentive, determined to be an optimal baseline incentive offer during the BPS:12/14 Phase 1 experiment with the calibration sample. Following the \$30 baseline offer, two different targeted interventions will be utilized for the BPS:12/17 responsive design approach:

- First Intervention (Incentive Boost): Targeted cases will be offered an additional \$45 over an individual’s baseline incentive amount. The \$45 amount is based on the amount identified as optimal during Phase 3 of the BPS:12/14 calibration experiment.
- Second Intervention (Abbreviated Interview): Targeted cases will be offered an abbreviated interview at 21 weeks (note that all cases will be offered abbreviated interview at 31 weeks).

Before each targeted intervention, predicted bias-likelihood values and composite propensity scores will be calculated for all interview nonrespondents. The product of the bias-likelihood and response propensity will be used to calculate the target importance score described above. Propensity scores above high and low cutoffs, determined by a review of the predicted distribution, will be excluded as potential targets during data collection⁶.

Pre-paid calibration experiment. It is widely accepted that survey response rates have been in decline in the last decade. Incentives, and in particular prepaid incentives, can often help maximize participation. BPS will test a prepaid incentive, delivered electronically in the form of a PayPal⁷ payment, to selected sample members. Prior to the start of full-scale data collection, 2,972 members of the previous respondent main sample will be identified to participate in a calibration study to evaluate the effectiveness of the pre-paid PayPal offer. At the conclusion of this randomized calibration study, NCES will meet with OMB to discuss the results of the experiment and to seek OMB approval through a change request for the pre-paid offer for the remaining nonrespondent sample. Half of the calibration sample will receive a \$10 pre-paid PayPal amount and an offer to receive another \$20 upon completion of the survey (\$30 total). The other half will receive an offer for \$30 upon completion of the survey with no pre-paid amount. At six weeks the response rates for the two approaches will be compared to determine if the pre-paid offer should be extended to the

⁶ These adjustments will help ensure that currently over-represented groups, high propensity/low importance cases, and very-difficult-to-convert nonrespondents are not included in the target set of nonrespondents. The number of targeted cases will be determined by BPS staff during the phased data collection and will be based on the overall and within sector distributions of importance scores.

⁷ A prepaid check will be mailed to sample members who request it. Sample members can also open a PayPal account when notified of the incentive. Any prepaid sample member who neither accepts the prepaid PayPal incentive nor check would receive the full incentive amount upon completion by the disbursement of their choice (i.e. check or PayPal).

main sample. For all monetary incentives, including prepayments, sample members have the option of receiving disbursements through PayPal or in the form of a check.

After the calibration experiment detailed above, the calibration sample will join with the main sample to continue data collection efforts. These are described in detail below, summarized in table 9, and their timeline is shown graphically in figure 1.

Table 9. Timeline for previous respondents

Phase	Start date		Activity	
	Calibration sample	Main sample	Calibration sample	Main sample
PR-1	Week 0	Week 7	Begin data collection; calibration sample for \$10 pre-paid offer versus no pre-paid offer	Begin data collection; make decision on implementation of pre-paid offer based on results of calibration
PR-2	Week 14	Week 14	Target treatment cases for incentive boost	Target treatment cases for incentive boost
PR-3	Week 21	Week 21	Target treatment cases for early abbreviated interview	Target treatment cases for early abbreviated interview
PR-4	Week 31	Week 31	Abbreviated interview for all remaining nonrespondents	Abbreviated interview for all remaining nonrespondents

Special data collection protocol for double nonrespondents. Approximately 3,280 sample members (group 4 in table 2) had sufficient information in NPSAS:12 to be classified as a NPSAS:12 study member but have neither responded to the NPSAS:12 student interview nor the BPS:12/14 student interview (henceforth referred to as *double nonrespondents*). In planning for the BPS:12/17 collection, we investigated characteristics known about this group, such as the distribution across sectors, our ability to locate them in prior rounds, and their estimated attendance and course-taking patterns using PETS:09. We found that while this group constitutes approximately 10 percent of the sample, 58 percent of double nonrespondents were enrolled within the private for-profit sectors in NPSAS:12. We found that over three-quarters of double nonrespondents had been contacted—yet had not responded. We also found, using a proxy from the BPS:04 cohort, that double nonrespondents differed by several characteristics of prime interest to BPS, such as postsecondary enrollment and coursetaking patterns. We concluded that double nonrespondents could contribute to nonresponse bias, particularly in the private for-profit sector.

While we were able to locate approximately three-quarters of these double nonrespondents in prior data collections, we do not know their reasons for refusing to participate. Without knowing the reasons for refusal, the optimal incentives are difficult to determine. In BPS:12/14, due to the design of the importance score which excluded the lowest propensity cases, nonrespondents who were the most difficult to convert were not included in intervention targeting. As a result, very few of the double nonrespondents were ever exposed to incentive boosts or early abbreviated interviews in an attempt to convert them. In fact, after examining BPS:12/14 data, we found that less than 0.1 percent were offered more than \$50 dollars and only 3.6 percent were offered more than \$30. Similarly, we do not know if a shortened abbreviated interview would improve response rates for this group. Therefore, we propose a calibration sample with an experimental design that evaluates the efficacy of additional incentive versus a shorter interview. The results of the experiment will inform the main sample.

Specifically, we propose fielding a calibration sample, consisting of 870 double nonrespondents, seven weeks ahead of the main sample to evaluate the two special data collection protocols for this hard-to-convert

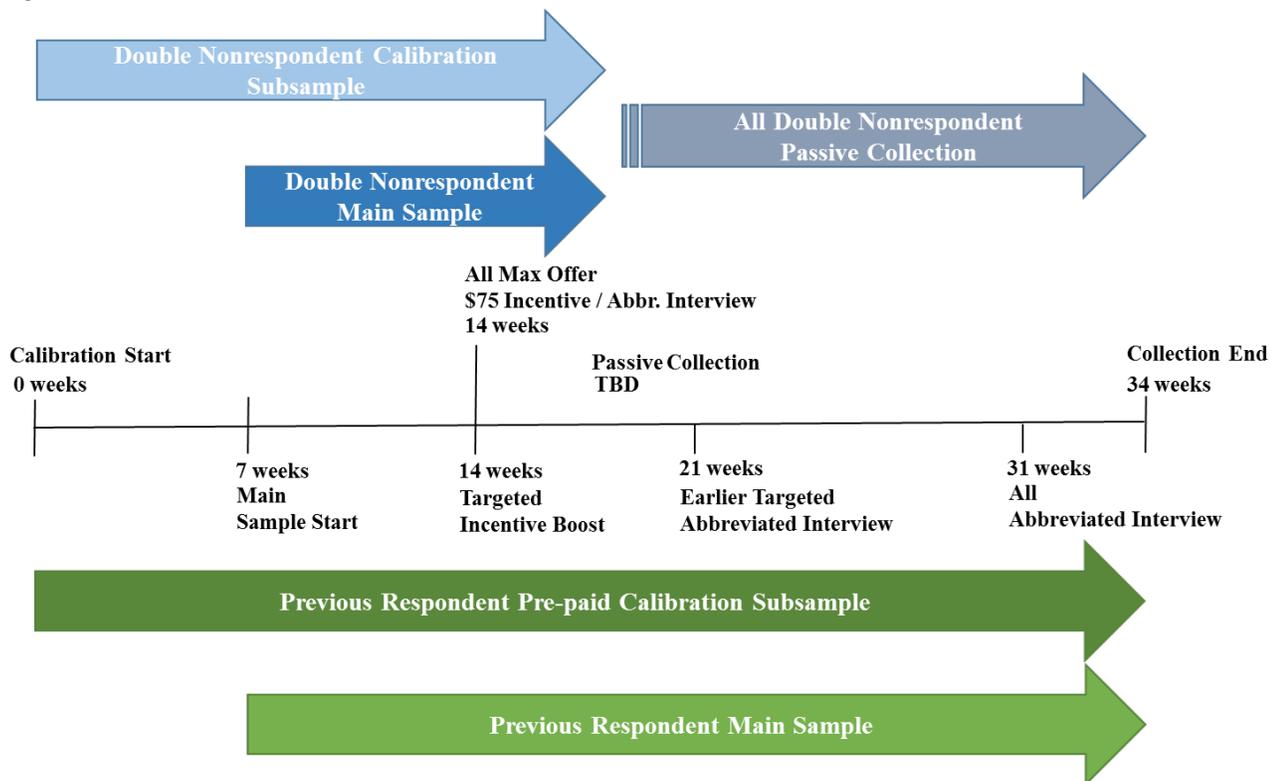
group: a shortened interview vs. a monetary incentive. A randomly-selected half of the calibration sample will be offered an abbreviated interview along with a \$10 pre-paid PayPal amount and an offer to receive another \$20 upon completion of the survey (\$30 total). The other half will be offered the full interview along with a \$10 pre-paid PayPal amount⁸ and an offer to receive another \$65 upon completion of the survey (\$75 total). At six weeks, the two approaches will be compared using a Pearson Chi-squared test to determine which results in the highest response rate from this hard-to-convert population and should be proposed for the main sample of double nonrespondents. If both perform equally, we will select the \$30 total baseline along with the abbreviated interview. Regardless of the selected protocol, at 14 weeks into data collection, all remaining nonrespondents in the double nonrespondent population will be offered the maximum special protocol intervention consisting of an abbreviated interview and \$65 upon completion of the interview to total \$75 along with the \$10 pre-paid offer. In addition, at a later phase of data collection, we will move this group to a passive status by discontinuing CATI operations and relying on email contacts. The timeline for double nonrespondents is summarized in table 10 and figure 1.

Table 10. Timeline for NPSAS:12 study member double interview nonrespondents

Phase	Start date		Activity	
	Calibration sample	Main sample	Calibration sample	Main sample
DNR-1	Week 0	Week 7	Begin data collection; calibration sample for baseline special protocol (full interview and \$75 total vs. abbreviated interview and \$30 total)	Begin data collection; baseline special protocol determined by calibration results (full interview and \$75 total vs. abbreviated interview and \$30 total)
DNR-2	Week 14	Week 14	Offer all remaining double nonrespondents \$75 incentive and abbreviated interview	Offer all remaining double nonrespondents \$75 incentive and abbreviated interview
DNR-3	Week TBD	Week TBD	Move to passive data collection efforts for all remaining nonrespondents; time determined bases on sample monitoring	Move to passive data collection efforts for all remaining nonrespondents; time determined bases on sample monitoring

⁸ All double nonrespondents will be offered the \$10 pre-paid PayPal amount in an attempt to convert them to respondents. For all monetary incentives, including prepayments, sample members have the option of receiving disbursements through PayPal or in the form of a check.

Figure 1. Data collection timeline



Evaluation of the BPS:12/17 Responsive Design Effort. The analysis plan is based upon two premises: (1) offering special interventions to some, targeted, sample members will increase participation in the aggregate for those sample members and (2) increasing participation among the targeted sample members will produce estimates with lower bias than if no targeting were implemented. In an effort to maximize the utility of this research, the analysis of the responsive design and its implementation will be described in a technical report that includes these two topics and their related hypotheses described below. We intend to examine these aspects of the BPS:12/17 responsive design and its implementation as follows:

1. *Evaluate the effectiveness of the calibration samples in identifying optimal intervention approaches to increase participation.*

A key component of the BPS:12/17 responsive design is the effectiveness of the changes in survey protocol for increasing participation. The two calibration experiments examine the impact of proposed features – a pre-paid PayPal offer for previous respondents and two special protocols for double nonrespondents.

Evaluation of the experiments with calibration samples will occur during data collection so that findings can be implemented in the main sample data collection. Approximately six weeks after the start of data collection for the calibration sample, response rates for the calibration *pre-paid offer* group versus the *no pre-paid offer* group for previous respondents will be compared using a Pearson chi-square test. Similarly the double nonrespondent group receiving the abbreviated interview plus the total \$30 offer will be compared to the group receiving the abbreviated interview plus the total \$75 offer.

2. *Evaluate sector group level models used to target cases for special interventions*

To maximize the effectiveness of the BPS:12/17 responsive design approach, targeted cases need to be associated with survey responses that are underrepresented among the respondents, and the targeted groups need to be large enough to change observed estimates. In addition to assessing model fit metrics and the effective identification of cases contributing to nonresponse bias for each of the models used in the importance score calculation, the distributions of the targeted cases will be reviewed for key variables,

overall and within sector, prior to identifying final targeted cases. Again, these key variables include base-year survey responses, institution characteristics, and sampling frame information as shown in table 7. During data collection, these reviews will help ensure that the cases most likely to decrease bias are targeted and that project resources are used efficiently. After data collection, similar summaries will be used to describe the composition of the targeted cases along dimensions of interest.

The importance score used to select targeted cases will be calculated based on both the nonresponse bias potential and on an *a priori* response propensity score. To evaluate how well the response propensity measure predicted actual response, we will compare the predicted response rates to observed response rates at the conclusion of data collection. These comparisons will be made at the sector group level as well as in aggregate.

3. *Evaluate the ability of the targeted interventions to reduce unit nonresponse bias through increased participation.*

To test the impact of the targeted interventions to reduce nonresponse bias versus not targeting for interventions, we require a set of similar cases that are held aside from the targeting process. A random subset of all sample members will be pulled aside as a control sample that is not eligible for intervention targeting. The remaining sample member cases will be referred to as the treatment sample and the targeting methods will be applied to that group. Sample members will be randomly assigned to the control group within each of the five sector groups. In all, the control group will be composed of 3,615 individuals (723 per sector group) who form nearly 11 percent of the total fielded sample.

For evaluation purposes, the targeted interventions will be the only difference between the control and treatment samples. Therefore both the control and treatment samples will consist of previous round respondents and double nonrespondents, and they will both be involved in the calibration samples and will both follow the same data collection timelines.

The frame, administrative, and prior-round data used in determining cases to target for unit nonresponse bias reduction can, in turn, be used to evaluate (1) unit nonresponse bias in the final estimates and (2) changes in unit nonresponse bias over the course of data collection. Unweighted and weighted (using design weights) estimates of absolute nonresponse bias will be computed for each variable used in the models:

$$|\bar{y}_r - \bar{y}_s|$$

where \bar{y}_r is the respondent mean and \bar{y}_s is the full sample mean. Bias estimates will be calculated separately for treatment and control groups and statistically compared under the hypothesis that treatment interventions yields estimates with lower bias.

BPS:12/17 Responsive Design Research Questions. With the assumption that increasing the rate of response among targeted, underrepresented cases will reduce nonresponse bias, the BPS:12/17 responsive design experiment will explore the following research questions which may be stated in terms of a null hypothesis as follows:

- Research question 1: Did the *a priori* response propensity model predict overall unweighted BPS:12/17 response?
 - H₀: At the end of data collection, there will be no association between *a priori* propensity predictions and observed response rates.
- Research question 2: Does one special protocol increase response rates for double nonrespondents versus the other protocol?
 - H₀: At the end of the double nonrespondent calibration sample, there will be no difference in response rates between a \$75 baseline offer along with a full interview and a \$30 baseline offer along with an abbreviated interview.

- Research question 3: Does a \$10 pre-paid PayPal offer increase early response rates?
 - H_0 : At the end of the previous respondent calibration sample, there will be no difference in response rates between cases that receive a \$10 pre-paid PayPal offer and those that do not.
- Research question 4: Are targeted respondents different from non-targeted respondents on key variables?
 - H_0 : Right before the two targeted interventions, and the end of data collection, there will be no difference between targeted respondents and non-targeted and never-targeted respondents in weighted or unweighted estimates of key variables not included in the importance score calculation.
- Research question 5: Did targeted cases respond at higher rates than non-targeted cases?
 - H_0 : At the end of the targeted interventions, and at the end of data collection, there will be no difference in weighted or unweighted response rates between the treatment sample and the control sample.
- Research question 6: Did conversion of targeted cases reduce unit nonresponse bias?
 - H_0 : At the end of data collection, there will be no difference in absolute nonresponse bias of key estimates between the treatment and control samples.

Power calculations. The first step in the power analysis was to determine the number of sample members to allocate to the control and treatment groups. For each of the five institution sector groups, 723 sample members will be randomly selected into the control group that will not be exposed to any targeted interventions. The remaining sample within each sector group will be assigned to the treatment group. We will then compare absolute measures of bias between the treatment and control groups under the hypothesis that the treatments, that is, the targeted interventions, reduce absolute bias. As we will be comparing absolute bias estimates, which range between zero and one, a power analysis was conducted using a one-sided, two-group chi-square test of equal proportions with unequal sample sizes in each group. The absolute bias estimates will be weighted and statistical comparisons will take into account the underlying BPS:12/17 sampling design; therefore, the power analysis assumes a relatively conservative design effect of 3. Table 11 shows the resulting power based on different assumptions for the base absolute bias estimates.

The final three columns of table 11 show how the power estimates vary depending upon the assumed values of the underlying absolute bias measures, and the third to last column specifically shows the worst case scenario where the bias measures are 50 percent. The overall control sample size is driven by sector group C which has the lowest available sample, and for some bias domains we may need to combine sectors C and D for analysis purposes. Given the sensitivity of power estimates to assumptions regarding the underlying treatment effect, there appears to be sufficient power to support the proposed calibration experiment across a wide range of possible scenarios.

Table 11. Power for control versus treatment comparisons across multiple assumptions

				Alpha	0.05	0.05	0.05
				Treatment Abs. Bias	0.4	0.2	0.125
				Control Abs. Bias	0.5	0.3	0.2
Sector Group	Sectors	Total Count	Control Sample	Treatment Sample	Unequal Group Power	Unequal Group Power	Unequal Group Power
A	1: Public less-than-2-year	10,202	723	9,479	0.916	0.966	0.925
	2: Public 2-year						
B	3: Public 4-year non-doctorate-granting	10,716	723	9,993	0.917	0.967	0.926
	4: Public 4-year doctorate-granting						
	5: Private nonprofit less than 4-year						
	6: Private nonprofit 4-year						

	nondoctorate						
	7: Private nonprofit 4-year doctorate-granting						
C	8: Private for-profit less-than-2-year	1,450	723	727	0.715	0.816	0.723
D	9: Private for-profit 2-year	2,989	723	2,266	0.861	0.933	0.873
E	10: Private for-profit 4-year	8,401	723	7,678	0.912	0.964	0.921

After the assignment of sample members to treatment and control groups, we will construct the two calibration samples: 1) previous respondents and 2) double nonrespondents. The calibration sample of previous respondents (n=2,972) will be randomly split into two equally sized groups of 1,486 sample members. One group will receive a \$10 pre-paid offer while the other will not receive the pre-paid offer. For a power of 0.80, a confidence level of 95 percent, and given the sample within each condition, the experiment of pre-paid amounts should detect a 5.0 percentage point difference in response rate using Pearson’s chi-square⁹. This power calculation assumes a two-sided test of proportions as we are uncertain of the effect of offering the pre-paid incentive. In addition, for the power calculation, an initial baseline response rate of 36.0 percent was selected given that this was the observed response rate after six weeks of data collection for the BPS:12/14 full scale study.

Similarly, we will randomly split the calibration sample of double nonrespondents (n=870) into two equally sized groups (n=435) where one group will receive a \$30 baseline offer with an abbreviated interview, while the other will be offered the full interview and a total of \$75. Using a power of 0.80, a confidence level of 95 percent, and given sample available within each condition, the experiment among double nonrespondents should detect a 5.0 percent point difference in response rate using Pearson’s chi-square. This power calculation assumes a two-sided test of proportions, as we have no prior data for which special protocol will perform better with respect to response rate. For this calculation, we assumed the six week response rate for the protocol with the lower response rate would be 5.0 percent. For a test between two proportions, the power to detect a difference is dependent on the initially assumed baseline response rate. In the proposed scenario, with a one-sided test, the baseline response rate would be the response rate for the approach with the lower response rate at six weeks. We assumed that this response rate would be 5% for power calculations. If the six week response rate differs from 5% then the power of the test to detect a 5% difference in response rates would fluctuate as shown in table 12 below.

Table 12. Power to detect 5 percent difference in response rate based on different baseline response rates

Alpha 0.05, Sample Size per group = 435, Detectable Difference 5%	
Assumed Response Rate	Power to Detect Difference
1%	98%
5%	80%
10%	61%

c. BPS:12 PETS Pilot Collection

The BPS:12/17 transcript collection provides an opportunity to test procedures and methods, particularly those associated with data that can be obtained from both interview and transcript sources. A pilot transcript collection, limited to respondents to the BPS:12/17 pilot test interview, offers an opportunity to compare transcript records with participants’ responses to interview questions about remedial and math course-taking. The results will inform development of interview questions and evaluation of what are the preferred sources for derived variables for the released data. The BPS:12/17 interview included questions about remedial courses, a separate question asking about completion of specific math topics (e.g. pre-college algebra or calculus), and about credits earned through pre-enrollment non-course activities (e.g. credit for military service or work experience). The transcripts collected for pilot test interview respondents will be reviewed and data will be collected exclusively for the analysis of these specific questions and for comparison of interview responses to data found on transcripts. Comprehensive transcript keying and coding will not be performed.

⁹ Calculated using SAS Proc Power. <https://support.sas.com/documentation/cdl/en/statugpower/61819/PDF/default/statugpower.pdf>

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

The study is being conducted by the National Center for Education Statistics (NCES), within the U.S. Department of Education. The following statisticians at NCES are responsible for the statistical aspects of the study: Dr. David Richards, Dr. Tracy Hunt-White, Mr. Ted Socha, Dr. Sean Simone, Dr. Elise Christopher, Dr. Gail Mulligan, Dr. Chris Chapman, and Dr. Marilyn Seastrom. NCES's prime contractor for BPS:12/17 is RTI International. The following staff at RTI are working on the statistical aspects of the study design: Mr. Jason Hill, Dr. David Wilson, Dr. Jennifer Wine, Mr. Darryl Cooney, Dr. Emilia Peytcheva, Ms. Nicole Ifill, Dr. T. Austin Lacy, Mr. Peter Siegel, and Dr. Andy Peytchev. Principal professional RTI staff, not listed above, who are assigned to the study include: Mr. Jeff Franklin, Mr. Michael Bryan, Ms. Kristin Dudley, Dr. Alexandria Radford, Ms. Donna Anderson, Ms. Chris Rasmussen, and Ms. Tiffany Mattox. Subcontractors include Coffey Consulting; HR Directions; Kforce Government Solutions, Inc.; Research Support Services; and Strategic Communications, Ltd.

C. References

- Becker, G.S. (1975). *Human Capital: A Theoretical and Empirical Analysis, With Special Reference to Education*. 2nd ed. New York: Columbia University Press.
- Chromy, J.R. (1979). Sequential Sample Selection Methods. In *Proceedings of the Section on Survey Research Methods, American Statistical Association* (pp. 401–406). Alexandria, VA: American Statistical Association.
- Folsom, R.E., Potter, F.J., & Williams, S.R. (1987). Notes on a Composite Size Measure for Self-Weighting Samples in Multiple Domains. *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, 792-796.
- Groves, Robert M.; Heeringa, Steven G. Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, July 2006, Volume 169, Issue 3, pages 439–457.
- Little, R. J. and Vartivarian, S. (2005). Does Weighting for Nonresponse Increase the Variance of Survey Means?. *Statistics Canada*, 31: 161-168.