

2012/14 BEGINNING POSTSECONDARY STUDENTS
LONGITUDINAL STUDY: (BPS:12/14)

Supporting Statement Parts B and C
Request for OMB Review
OMB # 1850-0631 v.8

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

October 24, 2013
Revised May 28, 2014

Table of Contents

- List of Tables.....ii
- B. Collection of Information Employing Statistical Methods.....3
 - 1. Respondent Universe and Sampling.....3
 - a. NPSAS:12 Full-scale Institution Universe and Sample.....3
 - b. NPSAS:12 Full-scale Student Universe and Sample.....5
 - c. Identification of FTBs in NPSAS:12.....6
 - d. BPS:12/14 Full-scale Sample.....7
 - 2. Procedures for the Collection of Information.....8
 - a. Statistical methodology for stratification and sample selection.....8
 - b. Estimation procedure.....8
 - c. Degree of accuracy needed for the purpose described in the justification.....8
 - d. Unusual problems requiring specialized sample procedures.....8
 - e. Any use of periodic (less frequent than annual) data collection cycles to reduce the burden.....8
 - 3. Methods for Maximizing Response Rates.....8
 - a. Locating.....8
 - b. Prompting Cases with Mail, E-mail, and SMS Contacts.....9
 - c. Telephone Interviewing and Help Desk Support.....10
 - d. Quality Control.....11
 - 4. Tests of Procedures and Methods.....11
 - a. Previous Studies.....11
 - b. BPS:12/14 Full Scale.....12
 - 5. Reviewing Statisticians and Individuals Responsible for Designing and Conducting the Study.....22
 - 6. Other Contractors’ Staff Responsible for Conducting the Study.....23
- C. Overview of Analysis Topics and Survey Items.....23
- References.....23

List of Tables

- Table 1. BPS:12/14 Sample Size by institution characteristics: 2012.....7
- Table 2. Summary of start dates and activities for each data collection phase, by sample.....17
- Table 3. Estimated Power, Within Decile, Using 10 percent and 25 percent Underlying Response Rates.....21
- Table 4. Estimated Power, Within Decile, Using 45 percent and 60 percent Underlying Response Rates.....21

B. Collection of Information Employing Statistical Methods

1. Respondent Universe and Sampling

The target population for the 2012/14 Beginning Postsecondary Students Longitudinal Study (BPS:12/14) full-scale 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 beginners (FTBs) from the 2011-12 National Postsecondary Student Aid Study (NPSAS:12) full-scale. Because the students in the BPS:12/14 full-scale sample come from the NPSAS:12 full-scale sample, this section also 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:12/14 full-scale sample is comprised of students from the NPSAS:12 full-scale sample who were determined to be FTBs, or were potential FTBs, as indicated by the NPSAS institution.

a. NPSAS:12 Full-scale Institution Universe and Sample

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.

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

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.

Postsecondary Education Data System (IPEDS) header, Institution Characteristics (IC), Fall and 12-Month Enrollment, and Completions files. All eligible students from sampled institutions comprised 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.

b. NPSAS:12 Full-scale Student Universe and Sample

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.

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.

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 FTBs, 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.

c. Identification of FTBs in NPSAS:12

Close attention was paid to accurately identify FTBs 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” FTBs) 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 to send for pre-sampling matching to administrative databases.

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. 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).

Prior to sampling, students over the age of 18 listed as potential FTBs 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 FTBs. Given that about 60 percent of FTBs receive some form of Title IV aid in their first year, this matching process could not exclude all listed FTBs 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 FTBs were matched to the National Student Clearinghouse (NSC) for further narrowing of potential FTBs 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 FTBs on the student sampling frame was sent to 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 among public 2-year and private for-profit institutions.

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.

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 FTBs 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 the study methodology report, publication number NCES 2006180, or BPS:04/06 methodology report, publication number NCES 2008184. 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.

d. 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 FTBs, and all will be included in the BPS:12/14 full-scale. In addition, the full-scale sample will include the 7,090 students who did not respond to the NPSAS:12 full-scale, but were potential FTBs 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 FTBs 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's are those NPSAS:12 non-respondents who appeared to be FTB's 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

2. Procedures for the Collection of Information

a. Statistical methodology for stratification and sample selection

The BPS:12/14 full-scale sample will consist of all interviewed and confirmed FTBs from the NPSAS:12 full-scale and a subsample of potential FTBs who did not respond to the NPSAS:12 full-scale. The sample size of 7,090 for the subsample was determined by estimating the impact on the unequal weighting effect due to subsampling and identifying the sample size that minimized this effect subject to available resources. The institution and student sampling strata used in the NPSAS:12 full-scale will be retained for use in analysis. The subsample of potential FTBs will be selected using stratified random sampling where the strata are defined by cross-classifying institution type and nonrespondent classification (study member or non-study member).

b. Estimation procedure

The students in the BPS:12/14 full-scale were also in the NPSAS:12 full-scale. Because the NPSAS:12 full-scale was a probability sample, sampling weights (which are the inverse of the selection probability) are available for each of the students in the NPSAS:12 full-scale and BPS:12/14 full-scale. These base weights, suitably adjusted for subsampling of NPSAS non-interview respondents, can be used in analyses of the results of the full-scale experiment. In addition, non-response weight adjustments and post-stratification adjustments will be applied to these base weights, and the fully adjusted weights used in the analysis of BPS:12/14 full-scale interview data.

c. Degree of accuracy needed for the purpose described in the justification

The sample will include interviewed FTBs from the NPSAS:12 full-scale and interview nonrespondents who were potential FTBs. Estimates of the overall unequal weighting effect due to subsampling indicate that the variance of overall estimates will increase by a factor of four under the proposed subsampling plan relative to the situation where no subsampling is performed and all interview nonrespondents are included in the BPS full-scale. The proposed subsample minimizes the overall impact on variance subject to resource considerations.

d. Unusual problems requiring specialized sample procedures

No special problems requiring specialized sample procedures have been identified.

e. Any use of periodic (less frequent than annual) data collection cycles to reduce the burden

BPS interviews are conducted no more frequently than every two years.

3. Methods for Maximizing Response Rates

a. Locating

Achieving a high response rate for BPS:12/14 depends on successfully locating sample members and gaining their cooperation. The availability, completeness, and accuracy of the locating data collected in the NPSAS:12 interview and institution records will impact the success of locating efforts. For BPS:12/14, all contact information previously collected for sample members, including current and prior address information, telephone numbers, and e-mail addresses, will be stored in a centralized locator

database. This database provides telephone interviewers and tracers with immediate access to all the contact information available for BPS:12/14 sample members and to new leads developed through locating efforts.

RTI's 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. RTI's proposed locating approach includes five basic stages:

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 for the BPS sample during NPSAS:12.
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:12/14 will also 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 several weeks prior to the start of data collection.

Panel Maintenance. Each year following the BPS:12/14 data collection, we contact sample members by mail and email to encourage them to update their contact information using a Web address update form. Conducting periodic panel maintenance activities allows us to maintain up-to-date contact information, increasing the likelihood of locating sample members in the future. Reducing the time elapsed between contacts maintains sample members' connection to the study, maximizing response rates for the subsequent follow-up. Receiving approval to conduct the panel maintenance activities as part of this OMB package will allow the activities to take place well in advance of the BPS:12/17 data collections.

The initial contact letter and email, the panel maintenance postcard and email, and the Web address update form may be found in Appendix F.

b. Prompting Cases with Mail, E-mail, and SMS Contacts

Past experience on recent postsecondary studies, including the BPS:12/14 field test, has shown that maintaining frequent contact with sample members with mailings, e-mails, and text messages sent at regular intervals will maximize response rates. The reminders to be sent regularly throughout the study include the following:

- Initial contact letter and e-mail sent to introduce the study and request updated contact information for the sample member.
- Data collection announcement letter and e-mail 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 e-mails 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.

Example contacting materials are also included in Appendix F.

c. Telephone Interviewing and Help Desk Support

Telephone interviewer training. Well-trained interviewers play a critical role in gaining sample member cooperation. The BPS:12/14 training program will include an iLearning module, which will provide staff with an introduction to the study prior to in-person training; 12 hours of in-person project training that will include training exercises, mock interviews, and hands-on practice; and ongoing training throughout data collection. The BPS:12/14 Telephone Interviewer manual will cover the background and purpose of BPS, instructions for providing Help Desk support, and procedures for administering the telephone interview.

Interviewing and Prompting. Telephone interviewing will be conducted using the same survey instrument as self-administered Web interviews. Data collection for the calibration study will begin in February 2013, and sample members selected for the calibration study will be encouraged to complete the survey on the Web during a 4-week early completion period. After four weeks, outbound calling will begin for sample members who did not participate during the early completion period. The remaining BPS:12/14 sample will begin data collection approximately 7 weeks after the calibration sample.

Telephone interviewing and prompting will be managed using the CATI Case Management System (CMS). The CATI-CMS is equipped to provide interviewers with access to all locating information and case histories, track scheduled appointments, and record notes about previous contacts with the sample member. Once cases are available for outbound calling, the CATI-CMS call scheduler will determine the order in which cases are to be called. The scheduler algorithm weighs several factors to determine the most appropriate number and time to call and automatically delivers the case to the interviewer for dialing. The CATI-CMS facilitates efficient case management by reducing supervisory and clerical time with integrated reports, monitoring appointments and callbacks automatically, and ensuring consistency of approach across interviewers.

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

Refusal Conversion. Avoiding refusals is important to maximizing the response rate for BPS:12/14. All interviewers will be trained in techniques that are designed to gain cooperation and avoid refusals whenever possible. Our training program stresses the importance of learning the most frequently asked questions, and when a sample member has concerns about participation, our interviewers are expected to respond to these questions knowledgeably and confidently.

When a refusal does occur, interviewers will enter comments into CATI to record all pertinent information about the case, including unusual circumstances or reasons given by the sample member for refusing. Through our monitoring efforts, we will identify interviewers who are especially adept at addressing sample members' concerns and will place them on a team of refusal conversion specialists. When a refusal is encountered, these specialists will review the call notes to determine the reason for the refusal and any concerns or questions posed by the sample member. The refusal specialists will determine whether a follow-up call is appropriate and will then tailor their response to resolve the issue.

d. Quality Control

RTI has developed quality control procedures based on "best practices" identified over the course of conducting thousands of survey research projects. Interviewer monitoring will be conducted using RTI's Quality Evaluation System (QUEST), a proprietary system that includes standardized monitoring protocols, performance measures, evaluation criteria, and reports to analyze performance data across interviewers.

As in previous studies, RTI will use QUEST to monitor approximately 7 percent of all recorded completed interviews, approximately 10 percent of recorded refusals, representing about 3 percent of overall budgeted interviewer hours. 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 (QC) 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.

4. Tests of Procedures and Methods

The design of the BPS:12/14 full-scale data collection—in particular, its use of responsive design principles to reduce bias associated with nonresponse—expands on data collection experiments designed for a series of preceding NCES studies. Although final results are not available for the most recent of those experiments, the design proposed for BPS:12/14 builds on what is known to date.

a. Previous Studies

NCES began efforts to reduce nonresponse bias through the use of responsive design principles in 2011, as it fielded the NPSAS:12 and B&B:08/12 field tests. In this early phase of work, the effort focused on the reduction of nonresponse bias through the identification of those cases least likely to

respond. As a result, two strands of work emerged: (1) the development of models that accurately predicted response propensity, and (2) the identification of incentive amounts that would convert potential non-respondents.

Although both the NPSAS:12 and B&B:08/12 field tests demonstrated that it was possible to develop models that were reasonably predictive of response propensity, results related to incentives were mixed. Taken together, results of experimentation in these studies suggested that relatively high incentive amounts (e.g., \$70 for B&B) are unlikely to convert cases with the lowest propensity scores, but that they can be useful in incentivizing response among cases with low-moderate to moderate scores. Additionally, differential response to reduced incentive amounts among high propensity cases in NPSAS:12 and B&B:08/12 suggest that the undergraduate population as a whole may be more sensitive to lower incentive amounts than students who have already completed their baccalaureate degree. However, neither study yielded evidence of bias reduction. Results from the ELS:2002/12 field test, which also focused on simple response propensity, produced similar, equivocal results.

As a result of its experience with the studies identified above, NCES sought to develop methods to better identify cases that were potentially bias-inducing. This approach involved: (1) identifying a set of substantive variables believed to be related to key study outcomes (typically frame variables for cross-sectional studies, but including prior-wave data for longitudinal studies), (2) using those variables to calculate the Mahalanobis distance of each case from either the full sample (for models estimated prior to data collection) or respondents (for models estimated during data collection), and (3) identifying ranges of distance scores for additional incentives or contractor effort. Generally, this “second generation” of responsive design efforts sought to replace targeting based on response propensity—which did not seem to reduce bias—with targeting based on a measure of multivariate difference. The conversion of so called “high-distance” cases, it was hypothesized, would reduce nonresponse bias.

To date, distance-based approaches have been employed in the B&B:08/12 full-scale, the ELS:2002/12 full scale, the BPS:12/14 field test, and the HSLs:09 2013 Update field test. While analyses of the B&B design and data collection for HSLs full-scale continue, the results from ELS and BPS FT are mixed. Although there was some evidence of a reduction in bias in the ELS and B&B designs, which involved a mix of increased contractor effort and higher incentives, no bias reduction was found in BPS, despite increasing response among high-distance cases through higher incentives.

b. BPS:12/14 Full Scale

After reviewing almost two years of related data, NCES experts are not convinced that they have identified an optimal model for characterizing a given case’s likelihood of contributing to nonresponse bias or identified an optimal procedure for delivering a given case’s ideal conversion intervention (typically an incentive). As a result, NCES proposes an experimental effort within the BPS:12/14 full scale that seeks to address the latter concern through an incentive calibration study and the former through the use of a two-component measure of “importance” that jointly considers a case’s likelihood of inducing bias and its propensity to respond. We begin by describing the proposed importance measure, proceed to a discussion of the calibration experiment, and then outline how the two will be combined in the BPS:12/14 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 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 high propensity cases can further bias estimates), and (3) they should be numerous enough to impact survey estimates within domains of interest. For example, targeting cases based on response propensities may reduce nonresponse bias, but it may fail if the targeted cases are extremely difficult to convert and do not respond to the intervention as desired.

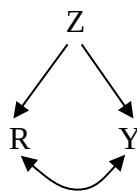
One approach to meeting these conditions is to target cases based on two dimensions: the likelihood of a case to contribute to nonresponse bias if not interviewed, and the likelihood that the case could be converted to a respondent. These dimensions form an importance score, such that:

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

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. Let i index sample members and j the data collection phases with responsive design interventions.

The importance score will be determined by the combination of two models: a bias-likelihood model and a response propensity model, which are described in detail below. The 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. There are a number of potential methods for combining these two models. Once in production, NCES will have more information about the distribution of both propensity and representation from the calibration study, which will allow it to explore linear and nonlinear functions that optimize the potential for nonresponse bias and available incentive resources.

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 the most direct way to reduce nonresponse bias in an estimate of a mean, for example.

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.

As needed in Phases 3-5, described below, we will conduct a logistic regression in order to predict the survey outcome (R) through the current phase using only substantive and demographic variables and their correlates from NPSAS:12 and the sampling frame (Z), and select two-way interactions. 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, school characteristics, and sampling frame information³. 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. Potential variables for the model include, but are not limited to:

- 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.

Response propensity (P(R)) model. Prior to the start of data collection, a response propensity model will be developed to predict likelihood to respond to BPS:12/14 based on NPSAS:12 data and response behavior. It will include variables from the base year that have been shown to predict survey response, including, but not limited to:

- responded during early completion period,
- interview mode (web/telephone),

³ Key variables will use imputed data to account for nonresponse in the base year data.

- ever refused,
- call count, and
- tracing/locating status (located/required intensive tracing).

In addition to estimating this initial model, we will evaluate the feasibility and potential benefit of adjusting the predicted response propensity scores at the start of each new phase based on response behavior in the previous wave.

The incentive calibration experiment. Prior to the start of full-scale data collection, approximately 10 percent of the main sample will be identified to participate in the incentive calibration study and will be randomly assigned incentive offers of \$0 to \$50, in \$5 increments. At the conclusion of this randomized calibration study, NCES will meet with OMB to discuss the results of the experiment and to determine the incentive amounts for the remaining 90 percent of the sample. NCES will calculate each calibration sample member's response propensity and, seven weeks prior to the main study, move the calibration sample to production. During production, response rates by incentive level within propensity decile will be monitored, with the goal of identifying the optimal incentive amount for each. Other than the experimental manipulation above, collection activities for the calibration sample will be identical to those used for the main study. These are described in more detail below, and summarized in table 2. Both main study and calibration samples will be exposed to a six-phase data collection protocol:

Phase 1: A four-week web only collection period. As their first phase of data collection begins, students in both BPS:12/14 subsamples – calibration and main study – will receive a letter asking them to log onto the web to complete the questionnaire. The calibration subsample will begin data collection on or around 2/18/2014 and the main subsample will begin data collection on or around 4/8/2014.

Baseline incentive amounts (ranging from \$0 to \$50 in \$5 increments) for the main sample will be determined by each case's predicted response propensity and the results of the calibration study. Results of the NPSAS:04 field test experiment showed that some students will respond when no incentive is offered (\$0). Although no a priori assessment of response propensity was estimated in the NPSAS:04 field test, cases that responded without being offered an incentive would have been considered high propensity cases. Given those results, and the findings from B&B:08/12 that showed there was not a significant difference in response rates between \$20 and \$35 dollar incentive amounts for the highest propensity cases, we expect that about 30 percent of sample members (about 11,000 cases) with the highest propensity to respond will participate when offered the lowest incentive amounts (between \$0 and \$20).

For the middle 40 percent, or about 13,000 cases, we expect that initial incentive offers at or above \$35 will be required to improve likelihood of response. In the B&B:08/12 field test this group was shown to be the most influenced by monetary incentives, but no comparable study has yet explored the potentially diminishing returns of larger monetary incentives. One goal of the calibration study is to better understand the relationship between response rates and monetary incentives, up to the proposed \$50 in phases 1 and 2 of data collection for the middle propensity group. For the remaining 30 percent who are in the low propensity group, the amount of the initial incentive will be set based on the results of the calibration study following consultations within NCES and with OMB.

Phase 2: Web and CATI data collection period. After the four-week, Phase 1 data collection period, in addition to the web interview option, outbound calling to sample members will begin and continue through the end of the survey. Promised incentives will be offered at the amount set at the start of data collection. This Phase 2 portion of data collection will continue for approximately 3 weeks.

Phase 3: First stage of a targeted nonresponse follow-up period. Approximately 7 weeks into data collection for each sample, predicted bias-likelihood values 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. Prior to creating the importance score, bias-likelihood estimates will be centered to the median predicted bias-likelihood score to ensure that no overrepresented cases are targeted. Additionally, propensity scores above high and low cutoffs, determined by a review of the predicted distribution, will be excluded as potential targets. These exclusions 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 that will be given special intervention in this phase. The number of targeted cases will be determined in conjunction with the COR, NCES, and OMB and by the overall and within sector distributions of importance scores.

We expect that about 58 percent of the initial sample (or 20,000 cases) still will be nonrespondents at this point in data collection. Among them, cases with the highest importance scores will be identified for special intervention during this phase. Note that those with the highest scores on both response propensity and bias likelihood will have the highest importance scores. Some cases that are high on either propensity or bias likelihood and low on the other measure may be considered in the group for special attention depending on their importance scores. The cut point for high importance scores will be guided by the results of the bias likelihood analysis. The cut point will be set in consultation with NCES and OMB. Those with low scores on both measures will have the lowest importance scores, and are not likely to be considered for special intervention in this phase. Targeted cases in the calibration subsample will be randomly selected to receive a \$15, \$30, or \$45 incentive increase over their assigned incentive level in Phase 1. We anticipate that about 30 percent of cases (about 6,500 cases) will be targeted in the main subsample to receive either a \$15, \$30, or \$45 incentive increase based on the results of the calibration study. All other cases will continue to be offered the baseline, promised incentive offered in Phase 1.

As a result of the two stages of the calibration study, although it is possible that some respondents could be offered up to \$95 in the total incentive, the total amount for each respondent will be driven by the respondents incentive that was set at the end of the first stage of the calibration study (i.e., some value between \$0 and \$50). Adding each of the three values to be tested in the second phase of the calibration study yields total incentive amounts of \$15 to \$95 in \$5 increments).

Phase 4: Second stage of the targeted nonresponse follow-up period. Approximately four weeks after the start of Phase 4 for each sample, predicted-likelihood values and importance scores will be recalculated for all remaining, nonresponding cases. Phase 4 targeted cases will be assigned to special field or call center staff who will use local exchanges to make outbound calls. Both targeted and non-targeted cases will be offered the same monetary incentives they were offered at the beginning of Phase 4.

Phase 5: Third stage of a targeted nonresponse follow-up period. Approximately 21 weeks into data collection for each sample, predicted-likelihood values and importance scores will be recalculated for any remaining nonresponding cases. Targeted cases in both the calibration and main subsamples will be

offered an abbreviated questionnaire that can be completed in approximately 10 minutes and on a mobile device. The incentive amounts determined in previous phases will continue to apply; nonrespondents newly targeted in Phase 5 will continue to be offered their original promised monetary incentive.

Phase 6: Abbreviated phase. Approximately 25 weeks into data collection for each sample, all remaining nonrespondent cases will be offered an abbreviated questionnaire that can be completed in approximately 10 minutes and on a mobile device. The offered incentives, as determined in previous phase(s), will continue to apply.

Table 2. Summary of start dates and activities for each data collection phase, by sample

Phase	Start date		Activity	
	Calibration subsample	Main subsample	Calibration subsample	Main subsample
1	2/18/2014	4/8/2014	Begin web collection; first calibration	Begin web collection; baseline incentives determined by results of first calibration
2	3/18/2014	5/6/2014	Begin CATI collection	Begin CATI collection
3	4/8/2014	5/27/2014	Target cases; second calibration	Target cases; incentive increase determined by results of second calibration
4	5/6/2014	6/24/2014	Field/local exchange calling for targeted cases	Field/local exchange calling for targeted cases
5	7/15/2014	9/2/2014	Abbreviated interview with mobile access for targeted cases	Abbreviated interview with mobile access for targeted cases
6	8/12/2014	9/30/2014	Abbreviated interview for all remaining nonrespondents	Abbreviated interview for all remaining nonrespondents

Analysis of the BPS:12/14 Full Scale Responsive Design Effort. Our analysis plan is based upon three premises: (1) sample cases that would contribute to nonresponse bias and have the potential to be converted to respondents can be identified at the beginning of the third and subsequent data collection phases, (2) the interventions during the third through fifth data collection phases are effective at increasing participation, and (3) increasing response rates among the targeted cases will reduce nonresponse bias by converting cases that would otherwise induce bias if they failed to respond. 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 the three topics and related hypotheses described below. We intend to examine these three aspects of the BPS:12/14 responsive design and its implementation as follows:

1. Evaluate model used to target cases for under-representativeness and potential impact

To maximize the effectiveness of the BPS:12/14 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. 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 was calculated based on both the non-response 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.

2. *Evaluate the effectiveness of each phase of data collection in increasing participation.*

The second key component of the BPS:12/14 responsive design is the effectiveness of the changes in survey protocol for increasing participation. Each phase introduces a feature – additional promised incentives, special field work, and an abbreviated mobile instrument. A calibration study will be used to determine the optimal baseline and Phase 3 intervention incentives.

Evaluation of the calibration study will occur during data collection so that findings can be implemented in the main subsample data collection. Approximately four weeks after the start of data collection for the calibration subsample, a logistic regression model will be estimated within each propensity decile. After estimating predicted response propensities, the minimum dollar amount associated with a predicted propensity that is not significantly different than the highest predicted propensity will be used as the baseline incentive for that propensity decile in the main subsample. Approximately four weeks after the start of the calibration study, a similar analysis will be conducted to determine the optimal incentive increase for the first phase of targeted nonrespondents.

In addition to analyzing data for the calibration study, we will evaluate response rates during the course of data collection, expecting an increase for targeted cases over the course of phases that include effective interventions. Furthermore, the cases that are not targeted with additional interventions will serve as a baseline for the pattern of responding over the course of the survey. Targeted nonrespondent completion rates are expected to increase with each phase, relative to the group that receives only standard recruitment efforts.

3. *Evaluate the ability to reduce nonresponse bias.*

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

$$\left| \frac{\bar{y}_r - \bar{y}_s}{\bar{y}_s} \right|$$

where \bar{y}_r is the respondent mean and \bar{y}_s is the full sample mean. The mean of these bias estimates can be tracked during the course of the survey and particular attention will be devoted to changes in the mean bias over each phase of data.

BPS:12/14 Responsive Design Research Questions. With the assumption that increasing the rate of response among targeted cases will reduce nonresponse bias, the BPS:12/14 responsive design experiment will explore the following research questions which build to our ultimate goal:

1. Did the a priori response propensity model predict BPS:12/14 response?
2. Are targeted respondents different from non-targeted respondents on key variables?
3. Do higher baseline incentive amounts increase early response rates?
4. Do higher additional incentive amounts increase response rates for targeted cases?
5. Did targeted cases respond at higher rates than non-targeted cases?
6. Did conversion of targeted cases reduce nonresponse bias?
7. Did conversion of high response propensity cases reduce nonresponse bias?
8. Did conversion of high bias likelihood cases reduce nonresponse bias?

Each of these questions may be stated in terms of a null hypothesis as follows:

Research question 1: Did the a priori response propensity model predict BPS:12/14 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: Are targeted respondents different from non-targeted respondents on key variables?

H₀: At the end of Phases 3, 4, and 5, 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 3: Do higher baseline incentive amounts increase early response rates?

H₀: At the end of the calibration sample Phase 1, there will be no difference in response rates by baseline incentive amount.

Research question 4: Do higher additional incentive amounts increase response rates for targeted cases?

H₀: At the end of the calibration sample Phases 3, 4, and 5, there will be no difference in response rates between targeted cases that receive additional incentive offers.

Research question 5: Did targeted cases respond at higher rates than non-targeted cases?

H₀: At the end of Phases 3, 4, and 5, and at the end of data collection, there will be no difference in weighted or unweighted response rates between targeted cases and non-targeted and never-targeted cases.

Research question 6: Did conversion of targeted cases reduce nonresponse bias?

H₀: At the end of Phases 3, 4, and 5, and at the end of data collection, there will be no difference in absolute nonresponse bias when targeted cases are included in analyses compared to when they are excluded.

Research question 7: Did conversion of high response propensity cases reduce nonresponse bias?

H_0 : At the end of Phases 3, 4, and 5, and at the end of data collection, there will be no difference in absolute nonresponse bias when targeted high response propensity cases are included in analyses compared to when they are excluded.

Research question 8: Did conversion of high bias likelihood cases reduce nonresponse bias?

H_0 : At the end of Phases 3, 4, and 5, and at the end of data collection, there will be no difference in absolute nonresponse bias when targeted high bias likelihood cases are included in analyses compared to when they are excluded.

Although not directly related to bias reduction, but relevant to the number of nonrespondents that ultimately require intervention and the nature of the Phase 3 intervention, are the questions:

1. What is the optimal baseline incentive amount within propensity deciles?
2. What is the optimal intervention amount for Phase 3 targeted nonrespondents?

These questions will be explored by estimating regression models and simulating marginal outcomes as outlined in King et al. (2000).

The calibration sample will be a random sample of approximately 10 percent of the 37,170 full-scale sample members. An estimated probability of response will be produced for each of the approximately 3,717 calibration sample members, and the calibration sample members will then be grouped into deciles constructed from the estimate response probabilities. The approximately 371 calibration sample members in each decile will be randomly assigned to one of eleven incentive levels; \$0, \$5, \$10, \$15, \$20, \$25, \$30, \$35, \$40, \$45, and \$50 so that approximately 33 calibration sample members will be assigned to each incentive level within each decile.

Calibration sample response rates will be examined, within each response probability decile, prior to the start of full-scale data collection in order to identify the incentive levels for use in the full-scale data collection. Formally, within each decile, ten cutpoints will be examined in order to determine the minimum incentive level that maximizes response rates for that decile. The first cutpoint will be set between the \$0 and \$5 incentive levels and the response rate for the \$0 incentive level will be compared with the response rate among individuals who received \$5 to \$50 as an incentive. The second cutpoint will be set between the \$5 and \$10 incentive levels, and the response rate for individuals who received \$5 or less will be compared with the response rate for individuals who received \$10 or more. The third through ninth cutpoints will be defined in a similar fashion, with the tenth cutpoint set between \$45 and \$50; the response rate for individuals offered the \$50 incentive will be compared with the response rate for individuals offered \$45 or less.

A power analysis was conducted in order to assess the ability to detect differences in response rates when considering any given cutpoint. For the purposes of estimating power, the total number of individuals within a given decile was assumed to be 363 (33 individuals for each of eleven incentive levels), an alpha level of 0.05 was used, and a one-sided test was specified. Given these assumptions, the power to detect any given difference still varies significantly depending upon the values of the proportions being estimated. In the tables below, sample power estimates are provided under different assumptions about the underlying response rates.

Using a two-group chi-square test of equal proportions with unequal sample sizes in each group, the power to detect a difference of 15 percent is 76 percent if the underlying true response rate is 25 percent for one group comprised of 33 individuals and the underlying true response rate is 10 percent for the other group comprised of 330 individuals. This power analysis corresponds to the comparison involving the first and tenth cutpoints because in those comparisons there will be two groups; one group comprised of 33 individuals and the other group comprised of the remaining 330 individuals. Additional power estimates were calculated for all cutpoints under the assumption that the underlying true response rates were 10 percent and 25 percent in the two groups. Table 3 shows the power estimates for the ten cutpoints with the lowest power of 65 percent occurring for the first cutpoint and the highest power of 98 percent occurring with the fifth and sixth cutpoints.

Table 3. Estimated power, within decile, using 10 percent and 25 percent underlying response rates

Cutpoint	Group 1 Sample Size	Group 2 Sample Size	Total	True Response Rate Group 1	True Response Rate Group 2	Power
1	33	330	363	0.1	0.25	65
2	66	297	363	0.1	0.25	89
3	99	264	363	0.1	0.25	96
4	132	231	363	0.1	0.25	97
5	165	198	363	0.1	0.25	98
6	198	165	363	0.1	0.25	98
7	231	132	363	0.1	0.25	97
8	264	99	363	0.1	0.25	95
9	297	66	363	0.1	0.25	91
10	330	33	363	0.1	0.25	76

Because the power estimates vary depending upon the assumed values of the underlying response rates, table 4 shows the power estimates at the lowest possible power, when the underlying true response rates approach 50 percent.

Table 4. Estimated power, within decile, using 45 percent and 60 percent underlying response rates

Cutpoint	Group 1 Sample Size	Group 2 Sample Size	Total	True Response Rate Group 1	True Response Rate Group 2	Power
1	33	330	363	0.45	0.60	50
2	66	297	363	0.45	0.60	71
3	99	264	363	0.45	0.60	82
4	132	231	363	0.45	0.60	86
5	165	198	363	0.45	0.60	88
6	198	165	363	0.45	0.60	88
7	231	132	363	0.45	0.60	86
8	264	99	363	0.45	0.60	82
9	297	66	363	0.45	0.60	71
10	330	33	363	0.45	0.60	50

Given the sensitivity of power estimates to assumptions regarding the underlying true response rates, there appears to be sufficient power to support the proposed calibration experiment across a wide range of possible true underlying response rates.

Calibration Study Results and Full-Scale Implementation

After approximately six weeks of data collection, response rates were compared across the 11 initial incentive levels for the calibration sample. Analyses were also conducted within propensity groups,

using 5 response propensity categories. For the three highest propensity groups combined, response rates were significantly higher with each additional \$5 incentive from \$0 to \$30, while no amount above \$30 resulted in a significantly higher rate. For the two lowest response propensity groups, the only statistically significant increase in response rate was found at the \$45 level. However, this finding was based on a small number of cases and was not supported by a linear trend in incentive amounts. NCES and OMB met to discuss these findings, and \$30 was implemented as the initial Phase 1 promised incentive for the remaining 90 percent of the full-scale sample.

The Phase 3 responsive design experiment tested three different additional incentive amounts on the calibration sample response rates and absolute nonresponse bias estimates. After approximately 7 weeks of data collection, cases were selected for targeting based on the importance score model. Nonrespondent cases that fell below the median on the bias likelihood score or that in either the highest 5 percent or lowest 15 percent of response propensities were not considered for additional targeting. From the pool of approximately 900 calibration sample nonrespondents eligible for targeting, the cases with the 500 highest importance scores (21 percent of the remaining calibration sample nonrespondents) were selected for the experiment. Cases were randomly assigned to three additional incentive treatment groups: \$0, \$25, or \$45. Analyses showed that \$45 was associated with significantly higher response rates, as well as a reduction in absolute bias across several key indicators. After consultation with OMB, NCES will use the importance model to select approximately 30 percent of the remaining sample nonrespondents at the start of Phase 3 and offer an additional \$45 incentive to those targeted cases.

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

Names of individuals consulted on statistical aspects of study design, along with their affiliation and telephone numbers are:

Name	Affiliation	Telephone
Dr. Jennifer Wine	RTI	(919) 541-6870
Dr. James Chromy	RTI	(919) 541-7019
Dr. Natasha Janson	RTI	(919) 316-3394
Mr. Peter Siegel	RTI	(919) 541-6348
Dr. David Wilson	RTI	(919) 541-6990
Dr. Bryan Shepherd	RTI	(919) 316-3482
Dr. Paul Biemer	RTI	(919) 541-6056
Dr. Emilia Peytcheva	RTI	(919) 541-7217
Dr. Andy Peytchev	RTI	(919) 485-5604
Dr. Alexandria Radford	RTI	(202) 600-4296
Dr. John Riccobono	RTI	(919) 541-7006

In addition to these statisticians and survey design experts, the following statisticians at NCES have also reviewed and approved the statistical aspects of the study: Dr. Tracy Hunt-White, Ted Socha, Dr. Matt Soldner, Dr. Sean Simone, and Dr. Sarah Crissey.

6. Other Contractors' Staff Responsible for Conducting the Study

The study is being conducted by the Sample Surveys Division of the National Center for Education Statistics (NCES), U.S. Department of Education. NCES's prime contractor is RTI. Principal professional staff of the contractors, not listed above, who are assigned to the study are provided below:

Name	Affiliation	Telephone
Ms. Nicole Ifill	RTI	(202) 600-4295
Mr. Jason Hill	RTI	(919) 541-6425
Ms. Donna Anderson	RTI	(919) 990-8399
Mr. Jeff Franklin	RTI	(919) 485-2614
Ms. Chris Rasmussen	RTI	(919) 541-6775
Ms. Tiffany Mattox	RTI	(919) 485-7791
Mr. Michael Bryan	RTI	(919) 541-7498
Dr. Jennie Woo	RTI	(510) 665-8276

C. Overview of Analysis Topics and Survey Items

The BPS:12/14 full-scale data collection instrument is presented in Appendix G. Many of the data elements to be used in BPS:12/14 appeared in the previously approved NPSAS:12 and BPS:04/09 interviews. New items, as well as items that are to be included in the abbreviated interview, are identified in Appendix G.

References

- 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.
- King, Gary, Michael Tomz, and Jason Wittenberg. 2000. Making the Most of Statistical Analyses: Improving Interpretation and Presentation. *American Journal of Political Science* 44: 341–355.
- Little, R. J. and Vartivarian, S. (2005). Does Weighting for Nonresponse Increase the Variance of Survey Means?. *Statistics Canada*, 31: 161-168.