MEMORANDUM

UNITED STATES DEPARTMENT OF EDUCATION
Institute of Education Sciences
National Center for Education Statistics

Date: March 11, 2011

To: Shelly Martinez, OMB

Through: Kashka Kubzdela, NCES

From: Tracy Hunt-White, NPSAS:12 Project Officer

Subject: Revisions to the NPSAS:12 Field Test Response Propensity Experiment Design

We would like to propose changes to the original design of the response propensity experiment planned for the NPSAS:12 field test data collection. The underlying premise for the approach has not changed: "In the NPSAS:12 field test, we plan to test a new methodology … that will minimize nonresponse bias by targeting cases that have a low likelihood of responding and a high likelihood of contributing to nonresponse bias" (OMB Supporting Statement, page 42). However, as we began examining the paradata that would be available for modeling during the early response phase of the field test data collection, and discussed how the original design could be implemented given that enrollment lists are received and sampled across several months on a flow basis, we realized that some modifications to the design were warranted.

Our proposed revisions to the design are discussed below.

First, in the initial design, we planned to identify high and low propensity cases at the completion of the early response phase, when students are contacted and invited to complete the survey online. We planned to categorize sample members as high or low propensity using (1) variables collected from enrollments lists provided by institutions prior to sampling (e.g., student level, level and control of the institution, first time beginner status); (2) variables collected from matches to the Central Processing System (CPS; e.g., full-time enrollment status, parents' education level); and (3) paradata collected during the early response phase. However, when in examining the results of preliminary modeling using the paradata from prior NPSAS data collections, we found that the paradata which would be available from the NPSAS early response phase were not contributing significantly to our ability to predict the likelihood of response. Consequently, a decision was made to base our propensity modeling solely on information provided by sources one and two.

Consequently, in order to set propensity levels for the NPSAS field test sample, RTI has examined NPSAS:04[1] sampling frame variables and the CPS variables to determine which variables are most predictive of response outcomes. Preliminary modeling of the data has already been completed on the 100,110 eligible cases in the NPSAS:04 full-scale data collection.

[1] Only NPSAS:04 data will be used since it included the sample of first time beginners as the NPSAS:12 sample will.

We found the following to be the variables that are predictive of a sample member's response propensity:

Table 1. Odds Ratios Obtained from NPSAS:04 Propensity Modeling using Logistic Regression

Variable	Odds Ratio	95% Odds Ratio Confidence Interval	
Student attends a 4-year institution	1.611	1.566	1.669
Student attends a less –than-2- year institution	.785	.744	.828
Student attends a public school	1.092	1.054	1.132
Student attends a private-for-profit school	.755	.716	.796
Student is a first-time-beginner	1.004	.975	1.034
Student is in a doctoral program	1.727	1.586	1.880
Student is an undergraduate	.829	.784	.875
Student is in school full-time	1.048	.998	1.100
Mother is a college graduate	1.070	1.026	1.117
Father is a college graduate	1.129	1.080	1.179
Student has missing CPS data	.625	.594	.658

Using the variables listed, the fit was significant (X^2 (11) = 4963.89, P <.0001). (The maximum pseudo-R^2 for the model was calculated to be 0.13. While low, this result is consistent with models of survey nonresponse and improves the predictability of response over chance.) We will assign a propensity score to students sampled for the NPSAS:12 field test using the set of variables listed above. Unlike the initial design, propensity scores will be assigned across institutions, rather than within institutions.

The second modification to the initial design of the response propensity experiment involves group assignment and the incentive offer. Because the propensity modeling will not rely on paradata from the early response phase, assignment to high and low propensity groups can occur at the time of sampling, rather than following the early response phase. The sample will be divided, first, into High and Low Propensity groups based on the results of the modeling. Within propensity groups, sample members will be assigned, at random, to an Experimental or a Control group (see table 2).

Table 2. Incentive Amounts to be Offered High and Low Propensity Sample Members, by Treatment Group

	Control	Experimental
High propensity	$30	$15
Low propensity	$30	$45

Among the High Propensity cases, the incentive offer for the Control group will be $30, an amount which has been used effectively in the last two NPSAS data collections. For the High Propensity Experimental group, the incentive offer will be $15. (If response rates in the High

Propensity Control and Experimental groups are shown to be equivalent, use of the lower incentive amount will help minimize the cost per completed interview for the data collection.) Sample members in the Low Propensity-Control group will also be offered the $30 incentive. Sample members in the Low Propensity-Experimental group will be offered a higher, $45 incentive.[2] The rationale for the different incentive amounts is that high propensity cases are already highly likely to participate and may do so irrespective of incentive amount. In contrast, cases in the low propensity group are not inclined to participate and, therefore, need the higher incentive amount to ensure their participation.

All other treatment of the low and high propensity groups, including locating activities and the number and types of contacts, will be the same. Because college students are often difficult to locate, an advantage of setting the incentive amount at the start of data collection is that, when we finally do locate them, they will receive the incentive message immediately. By waiting until the end of the early response phase to convey information about an incentive, we risk losing their attention and interest. Evaluation of the experiment will proceed as planned in the initial design, with the additional evaluation of the High Propensity group results. We will compare response rates among propensity groups, examine how well the propensity model predicted response outcomes, and investigate whether or not the low propensity group treatment was effective in minimizing response bias. We will also model propensity using the NPSAS:12 field test data to determine if there are other variables available from enrollment lists and/or CPS which can improve the predictive power of our original model.

If the propensity modeling is shown to successfully identify sample members who have high and low likelihood of response, a new incentive plan can be implemented for the full-scale data collection that maximizes the use of the project resources for data collection. Specifically, by motivating low propensity cases to respond with the strategic use of a higher incentive amount, data collection costs will be minimized since fewer resources will need to be invested to locate and interview the cases. At the same time, the low propensity cases will be more equally represented among the respondent groups, ultimately, reducing bias.

References

Peytchev, A., Riley, S., Rosen, J., & Lindblad, M. (2010). Reduction of nonresponse bias in surveys through case prioritization. *Survey Research Methods*, 4(1), 21-29.

[2]Approximately 2,265 sample members will be assigned to either the high or low propensity group. The propensity groups will be further split in half into the control and experimental groups with about 1,133 sample members in each group.