June 1, 2011

MEMORANDUM

To: Shelly Martinez, OMB

From: Elise Christopher, NCES

Through: Kashka Kubzdela, NCES

Re: Response to OMB Passback on ELS:2002/12 Field Test (OMB# 1850-0652)

May 24, 2011 Passback from OMB

1) Regarding the ELS incentive plans for the field test data collection, it does not appear that the main points from the lengthy conversations we've had with NCES and RTI on NPSAS and B&B are reflected here. This relates principally to two major emphases – first that we are not interested in raising the incentive level for high propensity cases over what those cases received last time. It perpetuates the impression that the real starting point for NCES and RTI is always to raise the amount from last time, and that response propensity is just an add-on related to response rates, despite the rhetoric on bias.

We appreciate your feedback and as part of our ongoing conversations with our colleagues on NPSAS:12 and B&B:09/12, we have revisited our respective plans. While there are some similarities in certain aspects of these plans, the ELS:2002/12 F3 field test incentive plan does not use the second follow-up incentive amount as the starting point for the F3 incentive. What a sample member received in the F2 survey does not dictate the incentive offered in the F3 field test (per exhibits A-2 and A-3 in Part A).

The plan for F3 is as follows:

- The cases will be split into a pre-data collection categorization of high propensity cases; low propensity early phase experiment cases; and low propensity control cases.
- During the 3-week early web period, all sample members will be offered the same amount: $25 (regardless of what they received for the F2 survey). This amount is lower than the $30 or $50 which was offered during the early phase of the F2 survey. (In F2, F1 non-respondents and "ever dropout" cases were offered the higher amount.) We believe that the $25 amount at this stage is appropriate and necessary, given the estimated burden (35 minutes – slightly longer than the actual F2 burden of 27 minutes) and population being surveyed (now in their mid-20s).
- Midway through this 3-week early web period, the cases in the low propensity early phase experiment group will receive telephone prompt to encourage completion of the web survey during the 3-week early period. These cases will continue to be offered the same $25 incentive amount.
- Once the 3-week period is finished, the 2nd (production) phase begins, which will last 6 weeks. It is at this point that we will re-determine our low- and high- propensity cases among remaining nonrespondents – and split low-propensity cases into experiment and control cases.
- For phase 2, the low propensity experiment cases will be offered $45 while the control cases and the high-propensity cases will continue to be offered $25.
- Phase 3 begins in week 10, and the incentive offer would be increased by $10 for these most difficult final cases: $55 for low-propensity experiment cases; $35 for the others. This compares with the F2 amounts of $60 and $40.

In short, with this plan, many sample members will be offered less than the amount they were offered in the F2 survey.

2) Second, we have very clearly expressed that NCES must demonstrate a relationship between substantive variables for which NCES is concerned about bias and the variables in the model. The variables for the model are mostly or all related to response propensity – how does increasing the response of a low propensity group directly address bias? This needs to be demonstrated and discussed not just assumed.

In the current environment of declining rates of survey participation, some concern has arisen that the inability to complete interviews with those who have a low propensity to respond is biasing survey estimates. From prior studies, it has been learned that low propensity cases are different than those with high propensity on certain key estimates. Including cases with the lowest propensity to respond could be a practical way to improve data quality in that it can reduce bias in the final survey estimates. For a nonresponse adjustment to be effective in reducing nonresponse bias, it needs to be associated with both a sample member's likelihood of participation and any survey variables of interest. Even then, however, these models are quite imperfect and attempts to obtain interviews from low propensity cases could be a superior supplemental approach to reducing nonresponse bias, for the reasons stated above. To be effective, an intervention during data collection aimed at reducing bias by targeting low propensity cases ideally (but not necessarily[1]) would meet two conditions: the response propensity should be a significant predictor of response outcome as well as being significantly associated with any survey variables of interest (observed only among respondents). These conditions were met in identifying cases for the ELS:2002 field test using information from prior rounds. The best fitting model found included the following variables as predictors: base year response status, F1 response status, F2 call count, whether the case has ever refused, whether some contact was ever made with the case, mother is a college graduate*, and whether the case has ever been in an AP class*. High propensity cases in the F2 FT had an overall response rate of 92%, while low propensity cases had a response rate of roughly 55%. While some survey variables were included in the response propensity model, it is informative to note that even some survey variables not ultimately retained as significant predictors do show an association to the calculated response propensities. This means that going after low propensity cases would bring respondents into the response pool *that could differ on many more key survey variables*. As shown below, calculated response propensities were significantly associated with certain important survey variables.

Table. ELS F2 FT - Correlations between Calculated Response Propensities and Selected Survey Variables

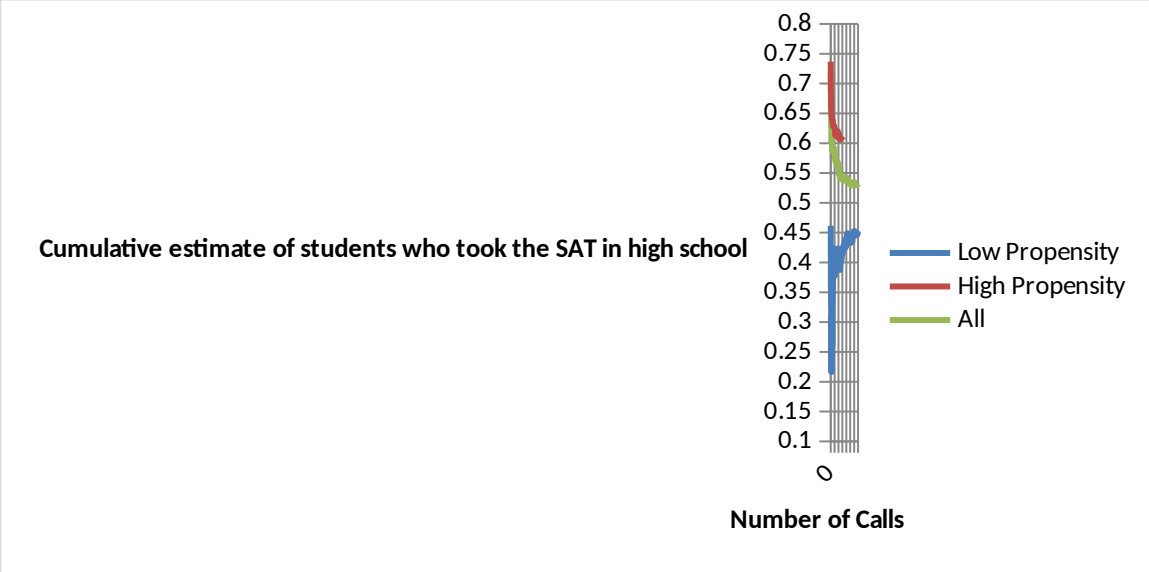
Survey Variable	Correlation between Variable and Calculated Response Propensity	Significance level	Estimate (overall)	Estimate (high propensity cases only)	Estimate (low propensity cases only)
Currently enrolled in postsecondary institution	0.16795	0.0003	.5401	.7503	.5270
Primarily a student (as opposed to a student and an employee or primarily an employee)	0.26110	<.0001	.2235	.3201	.1271

Another way to consider the potential effects of targeting low propensity cases is to view how estimates of key survey variables change with the inclusion of difficult to complete, low propensity cases. In the

[1] To the extent that the relationship between response propensity and the key survey variables is not the same among respondents and nonrespondents, only the first condition may be sufficient – estimated response propensity needs to be associated with the response outcome.

*Indicates this is a survey variable.

ELS:2002 F2 field test, respondents were asked if they had taken the SAT while in high school. The following graph shows the survey estimate for this item by level of effort (call counts). The Y axis is the cumulative estimate for the variable and the X axis is call counts, or a measure of level of effort.



After about 32 calls, the extra effort to complete low propensity cases has a demonstrated effect on the overall survey estimate. Even with this extra effort, the response rate for low propensity cases was significantly below that of high propensity cases. With this evidence that the estimate may change as a result of including low propensity cases, the ELS team believes it necessary to investigate the effects on data quality of bringing more low propensity cases into the response pool. Their values might further change the final estimate. One way to achieve this is through increasing incentives for low propensity cases to participate and directing more effort to these cases. The field test will experiment the use of higher incentives for low-propensity cases.

Analyses with the F3 field test data will be conducted to determine if the experiment resulted in a bias reduction associated with including more low-propensity cases in the ultimate response pool.

Analyses will also be conducted to determine how well the model predicted response propensity and whether the set of variables used to predict response propensity should be adjusted for the F3 full-scale study. The F3 field test should enable us to expand the pool of variables to build into the model for the F3 full-scale study since we will be able to identify additional F2 survey variables that best predict low- and high- response propensity for F3 response.

The approach we are testing in the ELS:2002/12 field test, as well as similar efforts that we are pursuing on other field tests of NCES studies (B&B:08/12 and NPSAS:12) will be used to refine our practices, relevant survey variables, and models for the full-scale study.

See additional analyses results below (from June 1, 2011).

May 26, 2011 Passback from OMB

3) What is the benefit of modeling based on past data if you are not going to use that information in practice except to make a phone call (the impact of which you are not proposing to separately analyzing, right)? Then this causes you to sort people into bins based on response behavior over a 3 week period (again, focusing on response behavior rather than bias, which seems like something you could have done in years past quite apart from propensity modeling. Separately, remember that NCSES (NSF) is experimenting with $0 and $30 for a similar population of low propensity cases, where the high propensity cases get nothing. So an assertion that the high group requires the same incentive (as they get for the first three weeks) may suggest that you don't really think you have people in the right bins to start.

Thank you for your additional comment. Upon further consideration, we have eliminated the early-phase prompting experiment from our plan so that the fidelity of our response-propensity higher-incentive experiment will not be compromised and such that it can be conducted for the entire data collection period.

The revised plan for the F3 field test is as follows:

- The cases will be split into a pre-data collection categorization of high propensity cases; low propensity experiment cases; and low propensity control cases.
- The low propensity experiment cases will be offered $45 while the control cases and the high-propensity cases will be offered $25.
- Beginning in week 10, the incentive offer will be increased by $10 for these most difficult final cases: $55 for low-propensity experiment cases; $35 for the others. This compares with the F2 amounts of $60 and $40.

The size and dictated yield of the F3 field test do not allow for additional testing of various incentive amounts. However, we are monitoring the NPSAS:12 field test propensity experiment and will continue to monitor NPSAS:12 and B&B:09/12 field tests to see if any aspects of their experiments may apply for the ELS:2002/12 full-scale study, and vice-versa. We also would welcome any information available from other studies (e.g., NCSES) as applicable.

Additional Analyses Provided to OMB on June 1, 2011

Using Response Propensity Models to Address Data Quality in ELS:2002 F3 Field Test

The response propensity approach under development at RTI is based upon several key assumptions that will be tested in the ELS F3 Field Test (FT). First, the approach rests on the assumption that low propensity cases, i.e. the cases least likely to respond to the survey, are fundamentally different from sample members with high response propensities. If differences between low and high propensity cases do exist and are large enough, survey estimates are likely to be affected. Thus, our second assumption is that low propensity cases contribute to nonresponse bias. With these assumptions in mind, the goal of the proposed approach is to first identify the cases with low propensity that we believe are likely to contribute to nonresponse bias, and then to increase their response propensity with a targeted intervention. To be effective, an intervention during data collection aimed at reducing bias by targeting low propensity cases ideally (but not necessarily[2]) would meet two conditions: 1) the calculated response propensity should be a significant predictor of response outcome and 2) response propensity should be significantly associated

[2] To the extent that the relationship between response propensity and the key survey variables is not the same among respondents and nonrespondents, only the first condition may be sufficient – estimated response propensity needs to be associated with the response outcome.

*Indicates this is a survey variable.

with survey variables of interest (observed only among respondents). The above assumptions have not yet been tested – the purpose of the proposed design is to allow us to test empirically with the field test sample whether nonresponse bias can be reduced by identifying and targeting cases with predicted low response propensity.

The ELS F3 FT will be used to empirically test whether or not *intervening* on low propensity cases can be a practical and effective method to improve overall survey estimates. The first step will be to determine if response propensities effectively predict survey response outcomes. It can then be determined whether the variance of the response propensity, because of an intervention, was lowered, and whether the association between the response propensity and any selected survey variables was reduced. A similar response propensity approach is being used (or has been proposed) for other studies conducted by RTI. RTI will carefully examine the results obtained in NPSAS, B&B, and HSLS along with the ELS F3 FT results. Based on extensive analyses of the field test experiments and careful examination of the ELS main study sample characteristics, plans will be developed for the respective full-scale collections.

Evidence from the ELS:2002 Second Follow-up Field Test

The ELS F2 FT can be used to illustrate how the approach outlined above could be effective in improving overall survey estimates. The best fitting response prediction model found in the ELS F2 FT included the following variables as predictors: base year response status, F1 response status, F2 call count, whether the case has ever refused, whether some contact was ever made with the case, mother is a college graduate*, and whether the case has ever been in an AP class*. High propensity cases in the F2 FT had an overall response rate of 92%, while low propensity cases had a response rate of roughly 55%. While some survey variables were included in the response propensity model, it is informative to note that even some survey variables not ultimately retained as significant predictors are correlated with response propensities, that is, there is a relationship between response propensity and substantive variables even if the substantive variables were not significant predictors of response propensity. For instance, in the ELS F2 FT, high and low propensity cases did show sometimes large differences across key survey variables. The table below shows 8 ELS F2 survey variables along with their reported overall estimate, the estimates separately for low and high propensity cases, as well as a significance test for that difference. Consider the variable current enrollment in a postsecondary institution which is a key ELS variable. The reported estimate for that variable among F2 FT respondents was .5401. However, as the table below shows, high and low propensity cases responded to this question very differently. And because our response among low propensity cases was far below that of high propensity cases, low propensity cases are probably underrepresented in the estimate for all interviewed F2 FT cases. It is likely that if additional low propensity cases would have participated, the overall estimate may have been different.

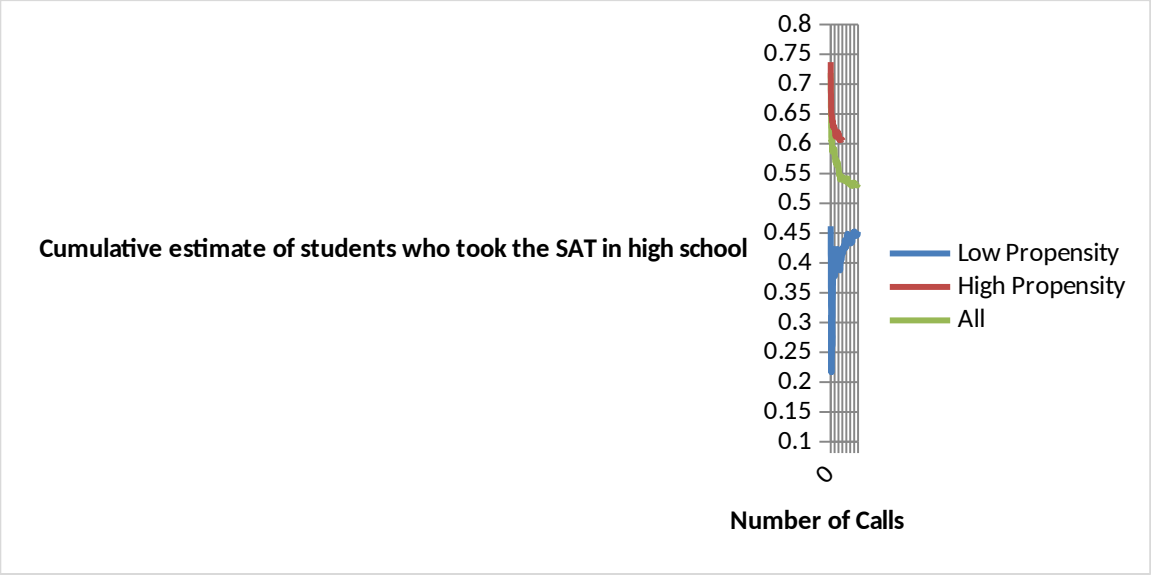
Table. ELS F2 FT – Estimates of Key F2 Survey Variables by Case Status

Key Variable	Estimate (overall)	Estimate (high propensity cases only)	Estimate (low propensity cases only)	χ^2	Significance Level
Currently enrolled in postsecondary institution	.5401	.6271	.4007	37.71	<.0001
Ever applied to postsecondary institution	.8510	.8791	.8150	6.05	.0139
Ever attended postsecondary	.8077	.8361	.7601	6.75	.0093

institution					
Took the SAT in high school	.5949	.6454	.5171	15.69	.0004
Took AP exams in high school	.2292	.2786	.1473	21.75	<.0001
Envisions a Bachelor's or higher	.6636	.6865	.6267	2.85	.0572
Makes payments on mortgage or rent	.2408	.2029	.3014	15.67	.0004
Contributes to the support of a dependent	.1911	.1762	.2192	18.04	<.0001

In the ELS F2FT, responses from low propensity cases did seem to matter for the overall accuracy of key survey estimates and many more survey variables could have been affected.

Another way to consider the potential effects of targeting low propensity cases is to examine how estimates of key survey variables change with the inclusion of difficult to complete, low propensity cases. In ELS F2 FT, respondents were asked if they had taken the SAT while in high school. The following graph shows the survey estimate for this item by level of effort to receive a response (call counts). The Y axis is the cumulative estimate for the variable and the X axis is call counts, or a measure of level of effort. The cumulative estimate is shown for all cases, and by propensity level.



After about 32 calls, the extra effort to complete low propensity cases has a demonstrated effect on the overall survey estimate. Even with this extra effort, the response rate for low propensity cases was significantly below that of high propensity cases. This example from ELS F2 FT suggests that: 1) low propensity cases are likely different in terms of how they respond to key survey items, and 2) intervening to bring low propensity cases into the response pool is sensible for overall data quality.

ELS F3 FT Response Propensity Incentive Experiment Design

The incentive plan for the ELS F3 field test is as follows:

- The cases will be split into a pre-data collection categorization of high propensity cases (50% of the F3 FT sample); low propensity experiment cases (25% of the sample); and low propensity control cases (25% of the sample).
- The low propensity experiment cases will be offered $45 while the control cases and the high-propensity cases will be offered $25.
- Beginning in week 10, the incentive offer will be increased by $10 for these most difficult final cases – the remaining nonrespondents: $55 for low-propensity experiment cases; $35 for the others. This compares with the F2 amounts of $60 and $40.

The size and dictated yield of the F3 field test do not allow for additional testing of various incentive amounts. (The designated total yield for the F3 FT is 500 cases.) Nonetheless, we are monitoring the NPSAS:12 field test propensity experiment and will continue to monitor NPSAS:12 and B&B:08/12 field tests to see if any aspects of their experiments may apply for the ELS:2002/12 full-scale study, and vice-versa.

Additional Analysis with the ELS Main Study Data

In parallel with the ELS F3 FT data collection, where experimental interventions are planned on cases with low predicted response propensities, we will conduct the above analyses with the ELS F2 main study data. These main study analyses, in addition to the F3 FT results, where we have intervened on low propensity cases during data collection, will provide a more complete picture of the potential utility of this approach as well as refine the model for the main study. For the main study, sample weights will be incorporated into these analyses and the model-building exercise.

Coordinating Activities with B&B and NPSAS

RTI is testing the response propensity approach to minimizing nonresponse bias in several field test studies for the first time. While there are variations in the study populations and in the "intervention" used in each, the goal of the approach is the same: identify cases with low response propensity, implement a targeted intervention to increase response propensity, and then evaluate the data to assess the impact to nonresponse bias. We plan to hold a meeting in the fall, after B&B, NPSAS, and ELS all finish their respective field test data collections, to review all propensity modeling experiment results, their relation to sample bias, and identify the best approaches for each study to yield the most benefits for data quality in full scale collections. The meeting is scheduled to piggy back on the B&B TRP meeting and will take place at the RTI-DC office on 11/15/2011. OMB will be invited and RTI will follow up to coordinate the meeting details.