

March 29, 2012

**MEMORANDUM**

**To:** Shelly Martinez, OMB  
**From:** Elise Christopher, NCES  
**Through:** Kashka Kubzdela, NCES  
**Re:** ELS:2002/12 Full Scale - Response to OMB Passback (OMB# 1850-0652 v.8)

---

**March 12, 2012 Passback from OMB**

- 1) The FAQs ask and answer “How will the information I provide on the Web remain private?” We would like NCES to be more precise in not mixing the concepts of privacy and confidentiality. The “privacy” of your answers should read either security, protection or confidentiality.

Response: We revised the text as follows:

**Q. How will the information I provide on the Web be protected?**

- A. All information entered through the ELS:2002 website requires users to log in with the matching study ID and unique password assigned to each person. The information you provide will be secured behind firewalls and will be encrypted during Internet transmission using Secure Sockets Layer (SSL) protocol.
- 2) The question asking for SSN includes as a reason future collections in the Appendix 1 version but that sentence is excluded in the Part C Survey Item Justification document. Please align the wording across the two documents, and if the sentence is retained, please explain why, since we understood this to be the last follow-up.

Response: The Appendix 1 reference to future collections has been deleted.

- 3) For the incentive plan, we think we understand how drop outs will be handled. However, we aren’t as sure about the rest of the sample. Please provide the “key metrics developed specifically for ELS:2002” and explain how they will be applied at each of the 3 checkpoints. In addition, please provide at least approximate targets of the number of cases and dollar amounts that would be applied at each of the three checkpoints. For example, will subsampling of those with distance scores above a cut-off be used to identify cases for more intensive follow up or will all cases in that group be handled the same way? And will the group be offered an amount between \$25 and \$55 after checkpoints 1 and 2, and the highest amount only after checkpoint 3? Please explain the logic for whatever criteria you plan to use.

Response:

For ELS:2002/12, we propose to use a Mahalanobis distance function calculated at three points in data collection to identify nonrespondent cases which are most unlike existing respondents.

The following variables will be used to calculate the Mahalanobis distance:

- Substantive data already collected/available from base year, first and second follow-up: enrollment status, parent's education, high school completion status, test score quartile, income quartile, native English speaker, school control, school locale, postsecondary enrollment, current work status, and whether the case was in school in grade 12.
- Paradata from current and prior rounds: response status from prior rounds, whether sample member contacted the help desk, whether sample member logged in but did not complete the questionnaire, number of contact attempts in the early data collection period, whether sample member made an appointment, and whether sample member told interviewer they would do the web interview. The Mahalanobis function incorporates a measure of the likelihood of ultimate response among current nonrespondents. Therefore, paradata are very important to include with the substantive data for factoring into the case-selection process thereby optimizing the selection itself.

Mahalanobis distance can be defined as the distance between a case and some group centroid. Therefore, cases with larger distance scores can be thought of as cases demonstrating large differences from the group. That is, these cases would be characterized by differences, for example, in their enrollment status, parent's education, high school completion status, etc.

Identifying these cases and presenting the specifically-targeted nonresponding cases with a higher incentive will in turn attempt to boost their participation and potentially reduce bias in estimates and also improve analytic power through higher sample sizes for these groups of cases of analytic interest.

Cases identified for targeting via their Mahalanobis distance (nonrespondent cases with distance scores furthest from the mean respondent) will receive a \$55 incentive – rather than the \$25 base incentive – in an attempt to encourage their participation.

We should note that ever-dropout cases are a critical set of cases and will receive \$55 from the start of data collection rather than being included in the Mahalanobis distance measure selection process during data collection. The approach with ever-dropouts is based on the particular difficulty gaining cooperation from this important subgroup as demonstrated during the F3 field test and prior rounds of data collection.

We propose to measure the distance functions at three points during data collection. At these points, the cases with the largest distance scores will be offered the \$55 incentive while the \$25 base incentive will remain intact for all other cases. Other activities that will be conducted to locate and interview targeted cases include performing pre-data collection intensive tracing, pursuing the cases in person with field locator/interviewers, and sending a \$5 prepaid incentive near the end of data collection. The case-identification for more-intensive effort will be done at three points during data collection.

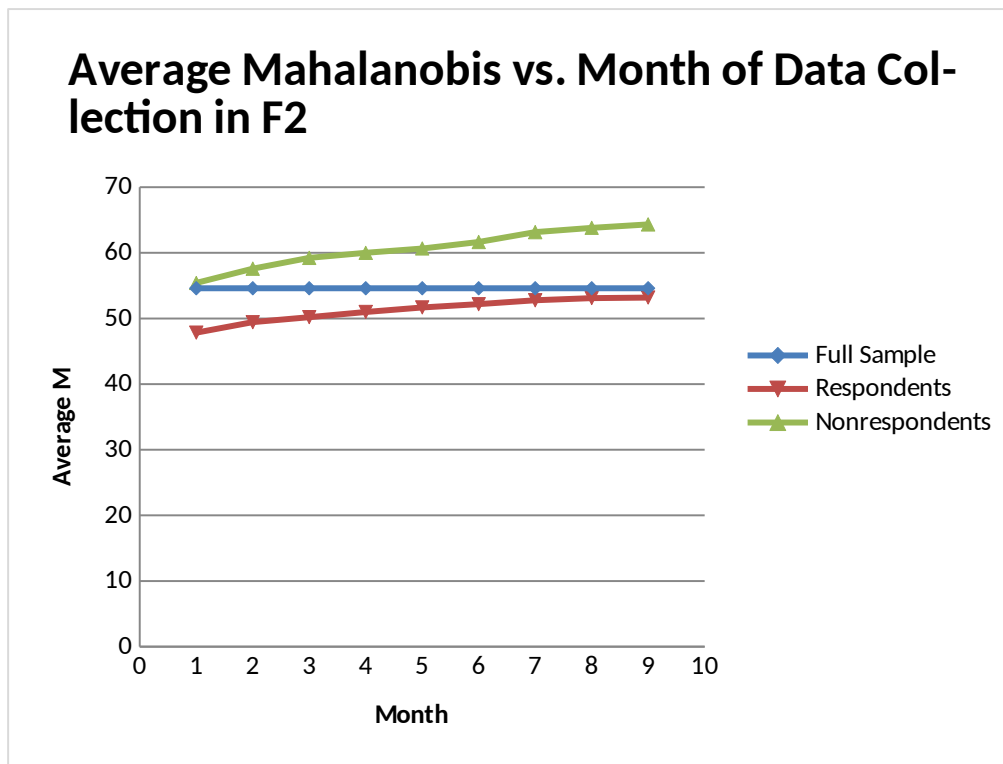
1. Our first target-case-selection point will be right before outbound CATI is scheduled to begin, 4 weeks into data collection.
2. The second point will be right before the CAPI period begins, 9 weeks into data collection.
3. The third point will be just prior to the pre-paid incentive period, approximately 8 weeks prior to the end of data collection.

At each juncture, the cases targeted will be those with the largest distance scores but not targeted in the prior phase(s). Also at the third and final point, cases identified for targeting will receive a \$5 prepaid incentive in addition to the \$55 incentive and other non-monetary activities, listed above. At this late point in data collection, these cases are likely to be the most critical for reducing the final mean distance function scores between respondents and nonrespondents, thus potentially improving analytic power in terms of sample size and reducing bias in estimates. We believe it prudent to be highly strategic with the

use of prepaid incentives, which are normally very effective. Again, cases not targeted will continue to be offered the \$25 incentive and will not be targeted for the more-intensive non-monetary activities.

Case targeting will be based on distance scores and anticipated yield. Based on the proportion of interviews projected across each of the phases of data collection, the distance scores will be segmented such that the following approximate numbers of cases are to be targeted at each cut point: 915 cases at the first cut point, 760 cases at the second cut point, and 830 at the third cut point. At each of these time points, we will evaluate the Mahalanobis values among the respondents and nonrespondents to identify the target cases.

We computed Mahalanobis values over time with the second follow-up data to identify variables for the third follow-up and to get a feel for how the metric will be implemented in the third follow-up. The figure below shows the Mahalanobis distance function calculated by month of data collection in the ELS:2002 second follow-up main study.



As shown in the figure, over time in the F2, the distance between respondents and nonrespondents grew slightly larger indicating that these groups may have become more dissimilar as data collection proceeded. The differences in mean distance scores between respondents and nonrespondents in F2 were not significant. However, this is illustrative of the potential benefits of carefully monitoring these values over time. If the differences were significant at the end, the risk of bias would have been increased. The goal for the ELS:2002 responsive design in the third follow-up main study will be to attempt to minimize the distances between respondents and nonrespondents over the course of data collection and therefore, reduce the risk of nonresponse bias.