Response to questions 9 and 10 of OMB Pass Back on BTLS

Thank you for bringing the issue of defining a criterion for success in the proposed incentive effectiveness experiment to our attention.

We did a power analysis using the sample sizes in the proposed study design to find the smallest differences in response rates that we could detect with alpha =.05 and alpha=.10 and beta=.8, assuming that we are trying to find differences between response rates that are on the order of 75 to 85 percent. We found that both alpha levels would only be able to detect fairly large differences (i.e., on the order of 7.5 to 17 percentage points. For example, we found that the smallest detectable differences between any two cells (in the range of interest, i.e., say, between 75% and 85%) in the proposed design would be 7.5 percentage points for a test between two cells of current teachers. Any other two-cell comparison would only detect larger differences.

We next considered a design that eliminated the no incentive cell and split the current teachers between the $10 and $20 incentive to see how much power we would gain for the tests for current teachers. With this design the former teachers and the nonresponding teachers could be distributed in any possible permutation, as there are not enough cases in those categories to be able to measure differences that are on the order of 5 to 10 percentage points.

With this second design, a comparison can be made between the $10 and $20 groups within current teachers that would detect a 5.5 percentage point difference (alpha=.05, beta=.8). The twenty-dollar incentive would be considered a success if a 5.5 percentage point or larger increase in response rate over that of the ten-dollar incentive group was observed. We concluded, however, that getting data on the impact of the incentive for current teachers, but not for former or nonresponding teachers, somewhat defeated the purpose of including teachers' status as an analytic strata in the analysis of the effectiveness of using different size incentives.

As a result of these investigations, we concluded that, given the existing sample sizes, it is not feasible to test for difference by teachers' current status classification. We next explored two alternatives. The first would be to divide the cases in each teacher status into the three incentive groups to insure a fair distribution of all teacher status categories across the three incentive levels, and then to collapse across the teacher status types. The test of the effectiveness of the incentive experiment would be measured by tests of the differences between the three incentive groups. This is likely to yield a more robust test of the increase in percentage points associated with each cash incentive relative to no incentive and of each cash incentive group relative to the other, but would not provide any insights into differences between cases in different teacher status categories.

The cases would be distributed as follows:

Teacher Status	No Incentive	$10 incentive	$20 incentive
Total	664	665	665

Note that each cell would include 505 current teachers, 101 nonresponding teachers, and 58 or 59 former teachers.

The three candidate tests of interest are:

- Test the no incentive group vs. the $10 incentive group for the total sample (n_1=664, n_2=665).
- Test the no incentive group vs. the $20 incentive group for the total sample (n_1=664, n_2=696).
- Test the $10 incentive group vs. the $20 incentive group for the total sample (n_1=665, n_2=665))

With an alpha of .05 and a beta of .8, this design would be expected to detect a difference of 6.5 percentage points between any two cells. With this design, the power of each of these tests to detect a 5 percentage point difference with an alpha of .05 is .56. Under the same testing conditions; their power to detect a 5.5 percentage point difference is .68. If the alpha is relaxed to .10, this design would be expected to detect a difference of 6.2 percentage points. With this design, the power of each of these tests to detect a 5 percentage point difference with an alpha of .10 is .68. Under the same testing conditions; their power to detect a 5.5 percentage point difference is .75.

As to a target to determine whether the increased incentive worked, we realized through the various power analyses performed that measuring any change below 5 percentage points was not feasible, given the available sample. The total cost of the incentive under this scenario is $19,930; however the marginal cost of the increase in the incentive from $10 to $20 is $6,650. Increasing the response rate 5 percentage points would yield an additional 33 cases at an added cost of $201.5 per case. It is difficult to quantify the reduction in error and the increase in quality of the data. However, note that with no experiments and a cash incentive of $10 for all cases, the cost of the incentive would be $19,940 or approximately the same cost that would be incurred with the no money, $10, and $20 experiment. We expect to learn from this whether the increased incentive will yield a significant increase in the response rate. This test would require an increase of just over 6 percentage points. IF this experiment works, we will able to put the information to good use in future waves of this data collection. However, there is an open question here as to whether the increase in responding cases in the $20 incentive group will offset the potential loss associated with the no incentive group.

The second alternative is a variation on the design just described, but it would provide more power to measure differences in the response rates between incentives of $10 vs. $20. Again the cases within each current teacher status category would be evenly distributed, but in this case they would be distributed into the two cash incentive groups.

The cases would be distributed as follows:

Teacher Status	$10 incentive	$20 incentive
Total	997	997

Note that each cell would include 577 or 578 current teachers, 151 or 152 nonresponding teachers, and 88 former teachers.

The test of interest is:

- Test the $10 incentive group vs. the $20 incentive group for the total sample (n_1=997, n_2=997))

With an alpha of .05 and a beta of .8, this design would be expected to detect a difference of 5.3 percentage points. With this design, the power of this test to detect a 5 percentage point difference with an alpha of .05 is .75. Under the same testing conditions; their power to detect a 5.5 percentage point difference is .84. If the alpha is relaxed to .10, this design would be expected to detect a difference of 4.7 percentage points. With this design, the power of the test to detect a 5 percentage point difference is .84. The power to detect a 5.5 percentage point difference is .91.

The total cost of the incentive under this scenario is $29,910; however the marginal cost of the increase in the incentive from $10 to $20 is $9,970. Increasing the response rate 5 percentage points would yield an additional 50 cases at an added cost of $199.4 per case. Thus the additional cost per case is approximately the same in the two experiments. As noted above, it is difficult to quantify the reduction in error and the increase in quality of the data; however, this experiment is more likely to allow us to detect a 5 percentage point difference because of the increased power associated with having two versus three experimental groups. IF this experiment is successful, it will yield more respondents than would be expected with a $10 incentive to all cases, or the no incentive, $10, and $20 experiment, and will provide information that is likely to result in an even larger response rate in the next wave if the $20 incentive is implemented. Although this alternative is more costly, and drops the $0 incentive, it is the option most likely to provide the data needed to detect a 5 percentage point difference between the two cash incentive levels (i.e., $10 vs. $20). We conclude that the incremental cost of $9,970 is well worth the expense.