

January 27, 2008

## **Late Stage Incentive Plans in the 2008 SESTAT Surveys**

### Background

NSF conducts three surveys of the science and engineering workforce: the National Survey of College Graduates (NSCG, OMB No. 3145-0141), the National Survey of Recent College Graduates (NSRCG, OMB No. 3145-0077) and the Survey of Doctorate Recipients (SDR, OMB No. 3145-0020). These surveys began data collection for the current round during the week of October 1, 2008. In July 2008, NSF had obtained OMB clearance for these surveys. For each of the surveys, NSF had proposed one or more monetary incentive experiments during the initial contact and follow-up portions of data collection to test the effectiveness of incentives in a variety of ways. NSF had also proposed that for each survey, an incentive would be offered to nonrespondents late in data collection, after the initial contacts and follow-up strategies had been exhausted. At the time of obtaining OMB clearance, NSF provided a very brief description of the approach for the design of this late stage incentive. The purpose of this document is for NSF to present a more detailed description of the late stage incentive plan design for each of the three SESTAT surveys.

### Overview

The overall strategy for the late stage incentive is to ensure that all sample members who have been subject to the standard survey data collection protocols and still remain as survey nonrespondents will have a probability of receiving a monetary incentive. In the plan we are proposing, a greater probability of selection for the incentive will be given to cases in those sampling cells where there are relatively lower response rates, in order to improve the accuracy of survey estimates (given that the sampling cells are aligned with the domains of interest for analysis).

To develop the strategy for designating which groups would have a greater or lesser probability of being offered an incentive, NSF prepared a table for each SESTAT surveys displaying for each 2008 sampling cell the 2006 final weighted response rate and the 2008 preliminary response rate as of December 31, 2008 (about two or three months into the data collection period). The sampling cells are used because it is at this level that NSF will make decisions about the incentive treatment. The 2006 final survey response rates are used as reference because they provide an indicator about the expected response rate for the sampling cells at the end of 2008 data collection. The 2008 preliminary response rates are shown so that we can make a determination of data collection progress to date.

In order to determine how to allocate its available limited resources for the monetary incentive to late stage survey nonrespondents, NSF will divide each survey's sampling cells into two groups: high and low probability of being offered a late stage incentive. The high treatment will be applied to cells where NSF would offer an incentive to a high proportion of members of that cell. The low treatment will be applied to cells where NSF would offer an incentive to a smaller proportion of the members of that cell. Among the "low" treatment group, 20% of that sample will receive the incentive; among the high treatment group, 100% will receive the incentive.

To define which sampling cells in each SESTAT survey would fit into the high and low treatments, NSF reviewed the 2006 final response rates, the 2008 preliminary response rates, and the difference between these two rates at this current stage in the 2008 data collection. Based on these data points, NSF developed the following broad parameters to use in making decisions for the incentive treatment.

1. **Parameter 1: High vs. Low Response Rate in 2006**  
A “low” response rate is any rate that is 90% or less of the overall 2006 final response rate. A “high” response rate is any rate that is higher than 90% of the overall 2006 response rate.
2. **Parameter 2: Fast vs. Slow Response Rate in 2008**  
A “slow” response rate is any response rate that is 80% or less of the overall 2008 preliminary response rate. A “fast” response rate is any rate higher than 80% of the overall 2008 preliminary response rate.
3. **Parameter 3: Big vs. Small Response Rate Difference Between 2006 and 2008**  
A “big” difference is any response rate difference that is 50% higher than the overall difference between the 2006 final and 2008 preliminary response rates. A “small” difference is any response rate difference that is less than 50% higher than the overall difference between the 2006 final and 2008 preliminary response rates. For example, in the 2006 NSCG, the final response rate was 87.5%, while it was 69.7% as of December 31, 2008, for an overall difference of 17.8%. Therefore, any cells that have a 2006-2008 difference of 26.7% or more (50% higher than 17.8%) will be labeled as cells with a “big” difference in current versus past response rate, while those with a difference of less than 26.7% will be labeled as cells with a “small” difference in current versus past response rate.

To stay within the budgetary restraints for each survey, given the response rates at the time when the incentives will be offered it may be necessary to adjust the cutpoints for the three parameters. However, the three parameters will remain the same as will the 20% of the low treatment group receiving the incentive.

Taking these three parameters together, a series of eight scenarios are possible. Table 1 provides the composition of possible incentive treatment groups and the rationale for the incentive treatment for each scenario. Table 2 provides the current response rate information for the surveys.

**Table 1: Response Rate Parameters and Assignment of Incentive Treatment**

<b>2006: High or Low Response Rate</b>	<b>2008: Fast or Slow Response Rate</b>	<b>2006- 2008: Big or Small Differ ence</b>	<b>Group Acro- nym</b>	<b>Late Stage Incentive Treatment</b>	<b>Rationale for Incentive Treatment</b>
Low	Slow	Small	LSS	High	Cells that did not perform well in 2006, and are still not performing well relative to other cells in 2008.
Low	Slow	Big	LSB	High	Cells that did not perform well in 2006, and are still not performing well relative to other cells in 2008.
Low	Fast	Small	LFS	Low	Cells that did not perform well in 2006, but are performing well relative to other cells in 2008.
Low	Fast	Big	LFB	High	Cells that did not perform well in 2006, and while they are performing better in 2008, still need to overcome a large difference to perform as well as 2006.
High	Slow	Small	HSS	Low	Cells that performed well in 2006, and although they are slow-performing in 2008, difference to overcome is small in order to perform as well as 2006.
High	Slow	Big	HSB	High	Cells that performed well in 2006, but are not performing well in 2008, and difference to overcome is large in order to perform as well as 2006.
High	Fast	Small	HFS	Low	Cells that performed well in 2006 and continuing to do so in 2008.
High	Fast	Big	HFB	Low	Cells that performed well in 2006 and 2008, even though the difference rate is large relative to other cells in 2008.

**Table 2: SESTAT Surveys Response Rates**

Survey	2006 Final Response Rate	2008 Preliminary Response Rate (as of 12/31/08)	Difference between 2006 and 2008 Response Rates
NSCG	87.5%	69.7%	17.8%
SDR	78.3%	55.5%	22.8%
NSRCG*	68.2%	32.5%	35.7%

\* NSRCG went into the field in 2008 about one month later than the other two surveys.

For each SESTAT survey, each sampling cell has been assigned a high or low incentive treatment status based on this plan. Table 3 shows the distribution of cases in the sampling cells by each group.

**Table 3: 2008 SESTAT Surveys Sampling Cells, Number of Cases and Response Rates by Incentive Treatment (as of 12/31/08)**

Group Acronym	Late Stage Incentive Treatment	2008 NSCG			2008 SDR			2008 NSRCG		
		No. of sampling cells	No. of non-respondent cases	2008 response rate	No. of sampling cells	No. of non-respondent cases	2008 response rate	No. of sampling cells	No. of non-respondent cases	2008 response rate
HFS	Low	132	17,395	71.1	75	10,577	61.5	134	7,495	36.1
LFS	Low	10	514	61.8	5	487	47.2	30	1,388	31.4
HFB	Low	11	649	59.9	5	180	47.8	3	178	33.1
HSS	Low	2	129	55.5	17	1,576	42.3	25	1,371	21.7
LSS	High	4	168	53.9	36	4,323	39.2	27	1,371	21.6
HSB	High	22	2,180	51.6	11	813	35.6	3	141	18.9
LSB	High	1	30	49.4	1	53	32.2	0	0	-
LFB	High	0	0	-	0	0	-	0	0	-
<b>Subtotal</b>	<b>Low</b>	155	18,687	70.8	102	12,820	59.7	192	10,432	34.7
<b>Subtotal</b>	<b>High</b>	27	2,378	51.8	48	5,189	38.8	30	1,512	23.8
<b>TOTAL</b>		182	21,065	69.7	150	18,009	55.5	222	11,944	32.5

### Incentive Costs

Each of the SESTAT surveys has an individual plan for amount and type of the late stage incentive. The NSCG will offer a \$20 prepaid incentive; the SDR will offer a \$30 prepaid incentive, and the NSRCG will offer a \$20 prepaid incentive.

Based on the status of the surveys as of December 31, 2008, the following proportions of sampling cells would receive the high or low late stage incentive treatments in each survey:

#### NSCG:

27 of the sampling cells would receive the high incentive treatment –  $100\% * 2,378 = 2,378$  cases

155 of the sampling cells would receive the low incentive treatment –  $20\% * 18,687 = 3,737$  cases

Total estimated incentive cost: \$122,300

39% of incentive cost to cases in high incentive treatment cells

61% of incentive cost to cases in low incentive treatment cells

SDR:

48 of the sampling cells would receive the high incentive treatment –  $100\% * 5,198 = 5,198$  cases

102 of the sampling cells would receive the low incentive treatment –  $20\% * 12,820 = 2,562$  cases

Total estimated incentive cost: \$232,530

67% of incentive cost to cases in high incentive treatment cells

33% of incentive cost to cases in low incentive treatment cells

NSRCG:

30 of the sampling cells would receive the high incentive treatment –  $100\% * 1,512 = 1,512$  cases

192 of the sampling cells would receive the low incentive treatment –  $20\% * 10,432 = 2,087$  cases

Total estimated incentive cost: \$71,980

42% of incentive cost to cases in high incentive treatment cells

58% of incentive cost to cases in low incentive treatment cells

## Sampling Parameters by Survey

The sample parameters for each survey are summarized below. Further details on the sampling cells are provided in the sampling documentation provided for each survey. Please note that the total number of sampling cells is smaller than the product of the levels of the sampling variables, because of some collapsing. For example, for some groups, male and female are collapsed for sampling efficiency.

NSCG:

There are 182 sampling cells, with four major stratification variables:

- 1) Demographic group (7 values) – this variable describes citizenship, disability status and race/ethnicity
- 2) Sex (2 values)
- 3) Degree (3 values) – this variable describes degree level
- 4) Field of degree (7 values) – this variable describes the field of highest degree

SDR:

There are 150 sampling cells, with three major stratification variables:

- 1) Demographic group (9 values) – this variable describes citizenship, disability status and race/ethnicity
- 2) Sex (2 values)
- 3) Field of degree (7 values) – this variable describes the field of doctorate

NSRCG:

There are 222 sampling cells, with four major stratification variables:

- 1) Demographic group (3 values) – this variable describes race/ethnicity
- 2) Sex (2 values)
- 3) Degree (2 values) – this variable describes degree level
- 4) Field of degree (17 values) – this variable describes the field of degree