

# Assessing Bias in Estimates in a Two Stage Design From an Early Close Out of the First Stage Data Collection: An Empirical Investigation Using NSRCG Sample Data

Donsig Jang and Xiaojing Lin, Mathematica Policy Research, Inc.

## Abstract

Reluctant respondents and low response rates have resulted in increased data collection costs to maintain the same level of response from one year to the next. Consequently, survey managers must assess the efficient allocation of a fixed budget to achieve the survey objectives when the survey is conducted. The list collection of college graduates from schools is a major component of the NSRCG first stage design and one that has considerable costs associated with it. In particular, data collection resources and time are concentrated on obtaining cooperation from a small set of late or ultimately non-responding schools. These resources could be better utilized elsewhere if the responses from schools can be obtained earlier. This paper focuses on the overall effect of school nonresponse if the list collection period was not extended and a higher school level nonresponse rate was accepted. With this objective, we assess the bias of survey estimates due to school-level nonresponse at varying response rates.

**Keywords:** Nonresponse analysis, NSRCG, School sample, Bias

## 1. Introduction

Reluctant respondents and low response rates have translated into increased data collection costs to maintain the same level of NSRCG response from one year to the next. Consequently, survey managers must assess the efficient allocation of a fixed budget to achieve the survey's objectives. Obtaining participant schools' cooperation is critical to constructing the sampling frame, which includes all graduates eligible for the survey. Clearly, the list collection of college graduates is a major component of the NSRCG design, with considerable cost implications.

---

**Acknowledgment:** Work on this article was supported and funded by the National Science Foundation, contact SRS-0244879.

**Disclaimer.** The views expressed are those of the authors and not necessarily those of the National Science Foundation

Person-level response rates have historically hovered around 80 percent but have recently dropped to less than 70 percent; at the same time, the school-level response rate has been almost perfect at about 99 percent (Wilson et al. 2005; Bandeh et al. 2007). However, achieving such high response rates comes with a substantial cost. In particular, data collection resources must be concentrated on a small set of late-responding schools, thus extending the data collection period and forcing the data collection contractor to devote considerable time and money to convincing reluctant schools to provide the requested lists of graduates. This chapter focuses on the effect of school nonresponse based on the assumption that the list collection period cannot be extended and that a higher school-level nonresponse rate would be acceptable. With this objective, we assess the bias of survey estimates attributable to school-level nonresponse at varying response rates.

## 2. Background and Outline of the Research

In recent years, it has become more challenging and expensive to obtain cooperation from sampled units—regardless of whether they are establishments or people. The NSRCG is no exception. Consequently, an important survey design issue is how to achieve survey objectives within a fixed budget. As mentioned, collecting the graduate lists for the NSRCG is associated with considerable cost. In particular, during the last few months of the collection period, resources must be concentrated on a small set of “difficult” schools; if schools responded more rapidly, resources could be used elsewhere. In addition, converting the nonresponding schools to respondents takes time; again, if schools responded more quickly, the graduate sample could be available earlier, permitting a shorter field collection period.

Our recent experience suggests a line of research that focuses on the extent to which the NSRCG school-level response rate might be affected by a shortened field collection period and the acceptance of a higher school-level nonresponse rate. With that in mind, we carried out the research outlined below to assess the extent of bias attributable to school-level nonresponse at varying response rates.

### 1. Data Sources

- 2003 NSRCG list collection status database

- 2003 NSRCG graduate sample frame
  - 2003 NSRCG survey response file
2. **Setting Up Three Response Rate Scenarios**
- Identify dates of each school file accepted by MPR
  - Sort the school by file-acceptance date
  - Treat the first XX percent of responding schools (early respondents) as respondents (four options:XX=85, 90, 95, 99 percent)
3. **Empirical Comparisons between Early- and Late-Responding Schools on Various Characteristics**
- Compare school characteristics between early- and late-responding schools
  - Compare demographic distributions of graduates from early- and late- responding schools
  - Compare person-level response rates of sampled graduates between early- and late-responding schools
  - Compare key NSRCG estimates between two groups of early- and late- responding schools before and after weighting adjustments

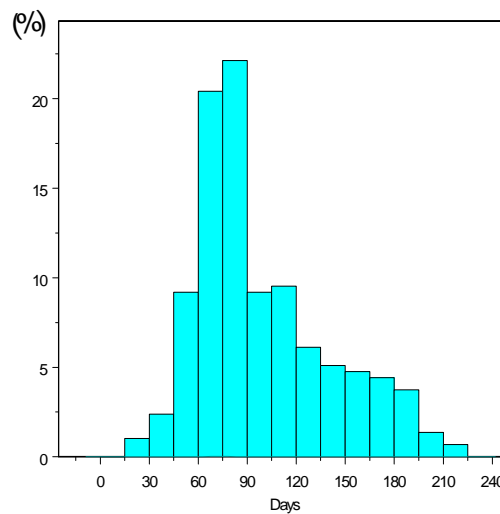
### 3. Empirical Investigations

For the 2003 NSRCG, MPR collected lists of graduates for a period of about seven months and achieved about a 99 percent response rate (only 4 refusals out of 300 schools selected for the 2003 NSRCG). We recorded the dates of acceptance of school-submitted lists during that period. Figure 1 shows the distribution of list submission dates for all 296 schools and indicates that the list collection period could have been shortened by about two months if the list collection response rate was compressed into a period yielding a 90 percent response rate. Shortening the list collection period by about a couple of months would allow more time for locating graduates and would save some resources for other list collection activities.

A primary concern is whether a shortened list collection period would adversely affect survey estimates. To respond to that critical concern, we empirically investigated the 2003 NSRCG data and executed a nonresponse analysis based on the following four options of response rates:

1. **Option 0:** 99 percent response rate (the current response rate)
2. **Option 1:** 95 percent response rate
3. **Option 2:** 90 percent response rate
4. **Option 3:** 85 percent response rate

**Figure 1: Distribution of List Submission Dates in 2003 NSRCG**



We first classified all 300 schools as “respondents” or “nonrespondents” based on their submission dates. For example, for option 1 (95 percent response rate), we treated the first 285 schools submitting lists as respondents and the remaining 15 schools as nonrespondents. Similarly, for option 2, we treated 30 schools as nonrespondents and, for option 3, 45 schools as nonrespondents. For the sake of convenience, we use the following notations to distinguish different response rate options and the corresponding responding/nonresponding groups.

- **Option 0:** Sample decomposition according to the 2003 final response status
  - EG0 consists of 296 responding schools
  - LG0 consists of 4 refusals
- **Option 1:** Sample decomposition based on 95 percent school-level response rate assumption
  - EG1 consists of 285 early-responding schools
  - LG1 consists of 11 late respondents and 4 refusals
- **Option 2:** Sample decomposition based on 90 percent school-level response rate assumption

- EG2 consists of 270 early-responding schools
- LG2 consists of 26 late respondents and 4 refusals
- **Option 3:** Sample decomposition based on 85 percent school-level response rate assumption
  - EG3 consists of 255 early-responding schools
  - LG3 consists of 41 late respondents and 4 refusals

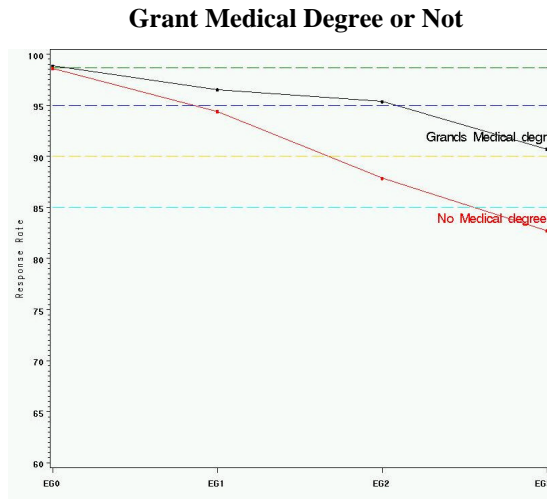
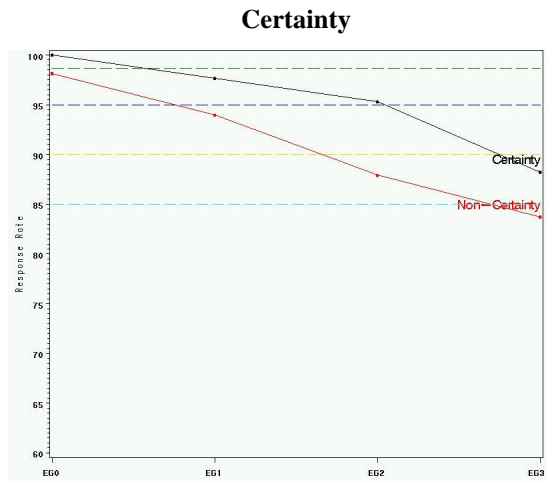
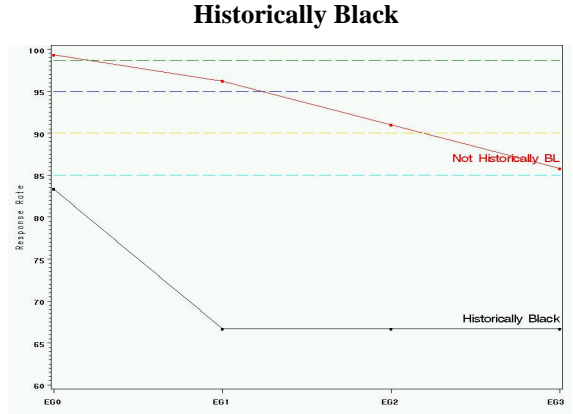
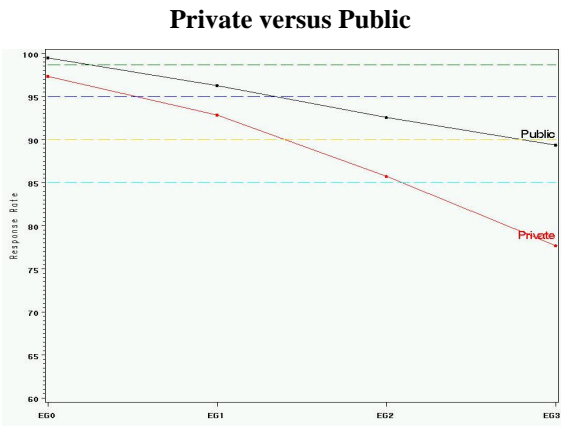
In the following sections, we present the results from empirical comparisons between early- and late-responding schools across the four response rate options.

### 3.1. School-Level Response Rates by School Characteristics

For each response rate option (0, 1, 2, 3), we compared school-level response rates by school characteristics such as control of school, whether the school is historically black, size of school (certainty versus noncertainty), and whether the school has a medical school (Figure 2). Some findings are summarized as follows:

- Private schools were less likely to respond early.
- Historically black schools were less likely to respond early.
- Schools large enough to be selected with certainty in the sample or granting medical degrees were more likely to respond early.

**Figure 2: School-Level Response Rates by Characteristics across Four Response Rate Options**



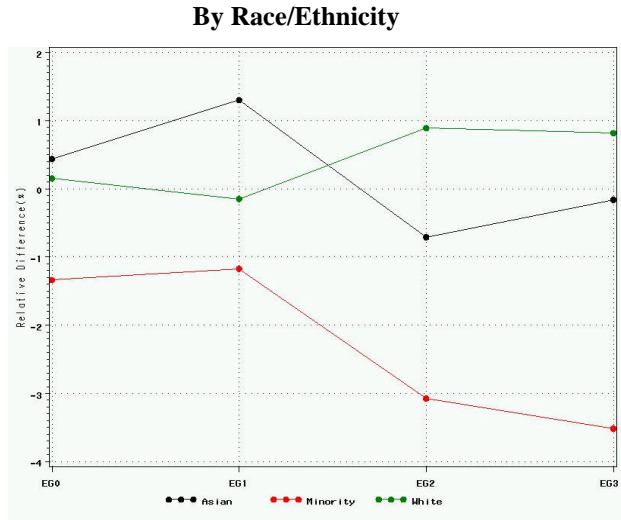
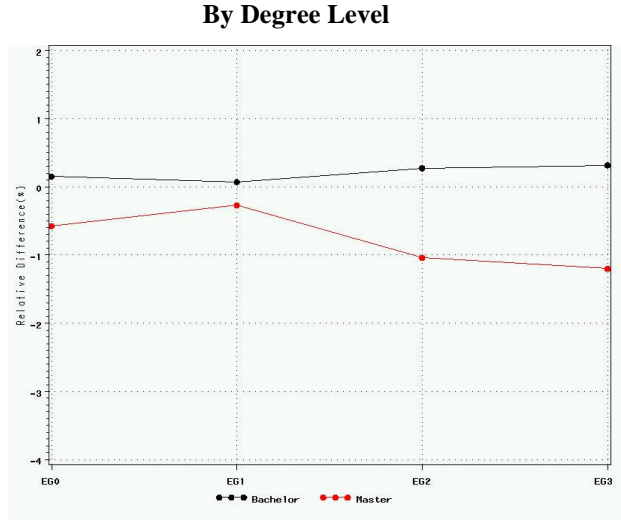
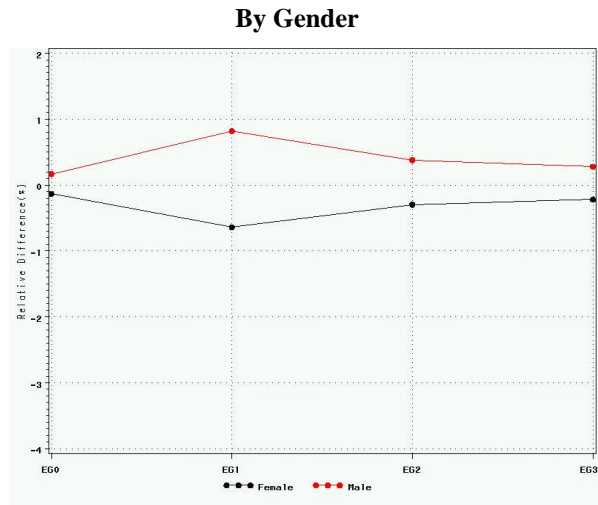
### 3.2. Demographic Composition of Graduates from Schools by School Response Status

We compared demographic distributions of graduates by schools' response status in order to identify any significant differences between early- and late-responding schools in terms of graduate characteristics. We calculated relative differences of proportions between responding schools and all sampled schools as follows:

$$RD(\hat{p}_{EGi}) = 100 \times \frac{\hat{p}_{EGi} - \hat{p}_s}{\hat{p}_s},$$

where  $\hat{p}_{EGi}$  is a weighted proportion estimate based on graduate counts from responding schools in  $EGi$ ,  $i = 0, 1, 2, 3$ ; and  $\hat{p}_s$  is a weighted proportion estimate based on graduate counts from all sampled schools. We calculated the weighted proportions based on the school-level sampling weights. Figure 3 shows the relative differences of key demographic proportions for each of four responding groups (EG0, EG1, EG2, EG3). A horizontal line at 0 may be used as a benchmark. The relative difference close to 0 means that the proportions by graduates' characteristic do not differ between respondents and the full sample.

**Figure 3: Relative Differences of Weighted Proportion of Demographic Groups by Their School-Level Response Status**



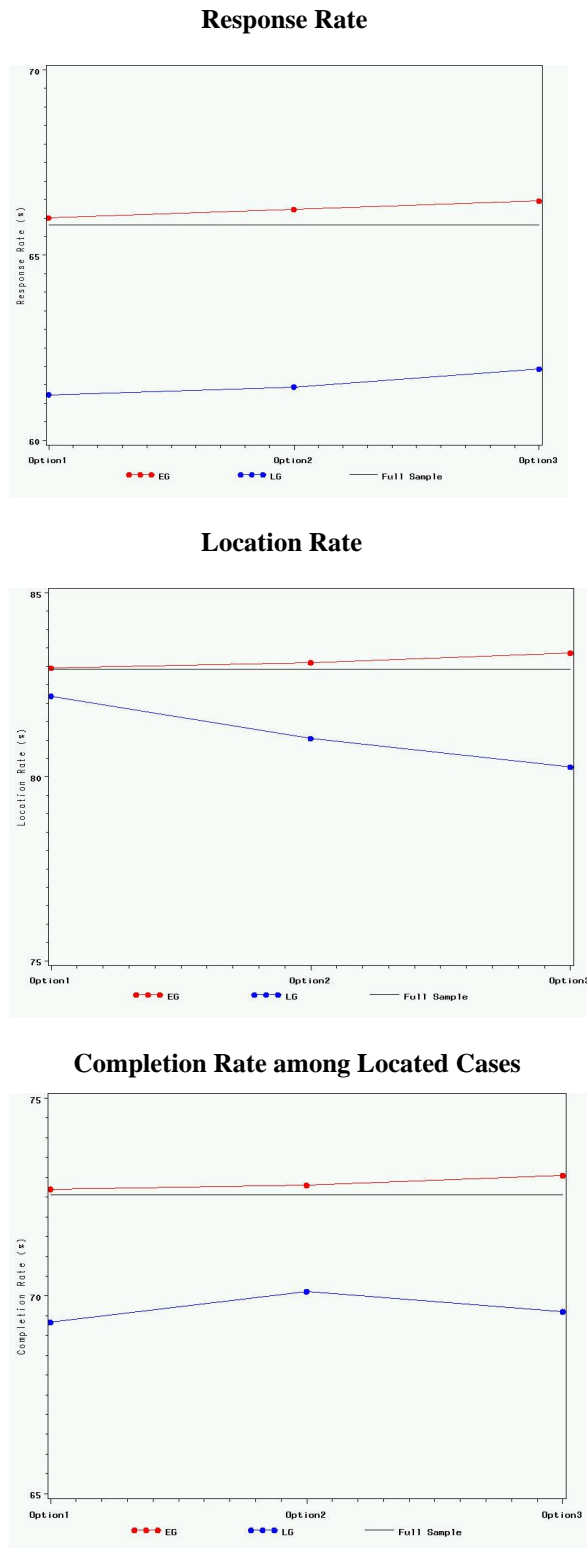
If the collection of graduate lists concluded before it was scheduled to conclude per the survey timeline, the sample could have underrepresented minority graduates, though not substantially; the relative difference of the sample's minority proportion decreases from about -1 percent to below 3 percent. This observation is consistent with the finding on school characteristics that historically black colleges were less likely to respond early.

### 3.3. Graduate-Level Response Rate Comparison between Early- and Late-Responding Schools

We compared response rates between graduates of early- and late-responding schools in order to determine if the response propensity of the sampled graduate might have depended on characteristics of the school from which the individual graduated. Figure 4 (first picture) presents response rates by key domains of graduates from early- and late-responding

schools (also full sample) under three response rate options (95, 90, 85 percent).

**Figure 4: Graduate-Level Response Rates by School Response Status**



The overall response rate for the 2003 NSRCG was 65.8 percent based on sampled units from all 296 responding schools. The rate increases slightly as the school-level response rate is compromised with fewer values. Specifically, the overall response rate increases to 66.0, 66.2, and 66.5 percent, respectively, as the school-level response rate is compromised as 95, 90, and 85 percent. Response rate differences become more evident if we directly compare response rates between graduates from early- and late-responding schools. With 95 percent of school-level response, graduate response rates were 66.0 and 61.2 percent, respectively, for the early- and late-responding schools. Similarly, the response rates are 66.2 and 61.5 percent, respectively, for early- and late-responding schools with a 90 percent school-level response rate and 66.5 and 61.9 percent, respectively, for early- and late-responding schools with an 85 percent school-level response rate.

Such strikingly different response rates between graduates from early- and late-responding schools are partly attributable to the unreliable locating information provided by late-responding schools reluctant to provide lists of graduates. Figure 4 (second picture) also shows location rates for each of the three response rate options and depicts a substantial difference for location rates between early- and late-responding schools. A more intensive, planned locating effort is suggested for future NSRCG list collection. However, it is also interesting to see a substantial difference in completion rates among located cases for early- and late-responding schools, a condition that is confounded with school characteristics. Late-responding schools were more likely to be minority-dominated schools, and minority graduates in turn were less likely to respond to the survey.

**3.4. Comparison of Key Survey Items between Early- and Late-Responding Schools**

To investigate whether graduates of early-responding schools are likely to exhibit characteristics different from those of graduates of late-responding schools on actual survey items, we compared estimates of key items such as degree level, employment status, salary, principal job, looking for work, and so forth and calculated relative differences of estimates between all respondents and respondents from each responding group (EG1, EG2, EG3) as follows:

$$RD(\hat{q}_{EGi}) = 100 \times \frac{\hat{q}_{EGi} - \hat{q}_{EG0}}{\hat{q}_{EG0}}$$

where  $\hat{q}_{EGi}$  is a survey estimate based on graduate respondent counts from responding schools in  $EGi$ ,  $i = 1, 2, 3$ ; and  $\hat{q}_{EG0}$  is a survey estimate based on all graduate respondents. Table 1 presents relative differences of estimates for several variables such as (1) looking for work, (2) principal job is S&E occupation, (3) principal job is S&E related health occupation, (4) principal job is S&E related non-health occupation, (5) principal job is non-S&E, (6) working for pay or profit, and (7) annual job.

For each survey item, we made comparisons by several domains, such as full sample, degree level, race/ethnicity, and gender. For most survey items, survey estimates from early-responding groups (EG1, EG2, EG3) do not seem to differ substantially from current survey estimates based on all survey respondents. We observed virtually no differences from the full sample based on estimates for all domains for the variables “working for pay or profit during the survey reference week” and “salary.” On the other hand, variables such as “looking for job among unemployed” showed noticeable differences between full sample-based estimates and early-responding group-based estimates. In particular, such differences become larger as school-level response rates decrease. For example, the master’s degree group exhibits a relative difference larger than an estimated 3 percent for “looking for job” for EG2. This difference becomes substantially larger (11 percent) for EG3 (i.e., school-level response rate is 85 percent). All other variables show moderate differences between responding groups.

It is worthwhile to mention that the differences discussed above were based on the current survey analysis weight. NSRCG weighting procedures are complicated and time-consuming; they account for (1) school-level selection probability, (2) school-level nonresponse, (3) graduate-level selection probability, (4) graduate-level nonresponse (separately for “not located” and “refusals”), (5) several degrees, (5) raking adjustment, (6) treatment of extreme weight, and (7) reraking. For details on NSRCG weighting, see Wilson et al. (2005). We replicated weighting procedures used for the full sample for each of the three data sets. After making the weighting adjustments, we noted that the observed differences in some subgroups became diluted, strongly indicating that the school-level response rate can be compromised down to 90 percent.

#### 4. Summary

We observed differences between early- and late-responding schools with respect to their characteristics and graduates’ demographic profiles:

- Private schools are less likely to respond early.
- Minority-dominated schools are less likely to respond early.
- Graduates of late-responding schools are less likely to respond than are graduates of early-responding schools.
- Person-level response rates may increase with compromised school-level response rates.
- Different data collection strategies can be considered for graduates of late- and early-responding schools.
- List submission dates can be used for weighting adjustment.
- The clustering effect may increase.
- Potential bias is observed for some survey items after dropping late-responding schools.
- For most survey items, differences were not substantial.
- Weighting may help reduce bias.

With this empirical investigation result, we will continue to achieve high response rates, such as 99 percent at the school level, but we will remain sufficiently flexible to compromise the school-level response rate down to 90 percent. To that end, we will perform real-time monitoring to check graduate counts by key domains and thereby assess any potential bias before stopping list collection below a response rate of 99 percent.

#### REFERENCES

- Bandeh, L., D. Jang, N. Duda, M. Satake, M. Bozylinsky, H. Haixia, and L. Xiaojing. “2006 National Survey of Recent College Graduates: Sampling Frame Development, Sampling and Location Procedures.” Princeton, NJ: Mathematica Policy Research, Inc., April, 2007.
- Wilson, C., D. Jang, T. Barton, M. Pierzchala, K. Kang, and J. Tsapogas. “2003 National Survey of Recent College Graduates: Methodology Report.” Washington, DC: Mathematica Policy Research, Inc., November 2005.

**Table 1: Relative Difference Comparison Before and After Reweighting for Key Survey Items**

Items	Domain	Final weight			Adjusted weight		
		285 schools	270 schools	255 schools	285 schools	270 schools	255 schools
Looking for work (LOOKWK_I)	ALL	-1.50%	-2.84%	-1.32%	-1.80%	-2.44%	0.51%
	Bachelor	-1.15%	-2.54%	1.12%	-1.58%	-2.75%	1.96%
	Master	-2.84%	-3.97%	-10.69%	-2.69%	-1.24%	-5.20%
	White	2.10%	1.65%	4.42%	1.59%	1.59%	4.25%
	Asian	-5.79%	-5.96%	-4.24%	-5.63%	-5.86%	-0.52%
	Minority	-1.66%	-5.54%	-6.26%	-1.74%	-4.20%	-3.43%
	Male	-1.31%	-6.25%	-6.10%	-1.23%	-5.04%	-4.91%
	Female	-1.68%	0.07%	2.74%	-2.28%	-0.24%	5.11%
Principal Job is S&E occupation (OCCUP_1_I)	ALL	0.01%	-2.20%	-1.45%	-0.10%	-0.04%	1.03%
	Bachelor	0.05%	-2.45%	-1.28%	0.04%	0.26%	2.38%
	Master	-0.42%	-2.14%	-2.44%	-0.39%	-0.60%	-1.88%
	White	0.99%	0.78%	2.18%	0.55%	2.73%	4.46%
	Asian	-1.31%	-5.16%	-4.77%	-0.90%	-3.02%	-1.53%
	Minority	-1.17%	-4.87%	-3.77%	-1.12%	-2.56%	-1.95%
	Male	-0.75%	-2.29%	-0.30%	-0.60%	0.28%	3.17%
	Female	0.93%	-1.55%	-2.77%	0.62%	-0.41%	-1.96%
Principal Job is S&E health-related occupation (OCCUP_2_I)	ALL	1.97%	1.57%	2.13%	0.57%	1.62%	1.60%
	Bachelor	1.93%	2.82%	3.52%	0.30%	1.55%	1.85%
	Master	1.72%	-1.59%	-1.48%	1.19%	1.88%	1.13%
	White	1.98%	1.04%	0.27%	0.16%	-0.38%	-0.72%
	Asian	1.70%	2.65%	6.20%	0.97%	3.90%	6.99%
	Minority	0.92%	1.26%	-0.08%	-0.58%	1.38%	-2.31%
	Male	2.01%	2.71%	2.94%	0.85%	1.52%	1.34%
	Female	1.31%	0.27%	0.27%	-0.20%	2.32%	2.02%
Principal Job is S&E- related non-health occupation (OCCUP_3_I)	ALL	-1.18%	1.62%	1.54%	0.02%	0.56%	1.28%
	Bachelor	-1.50%	0.67%	-0.19%	0.08%	0.51%	0.75%
	Master	-0.50%	3.91%	5.88%	-0.13%	0.77%	2.79%
	White	-1.21%	0.94%	-0.26%	-0.25%	0.33%	-0.23%
	Asian	-0.85%	2.23%	3.48%	0.31%	0.45%	2.22%
	Minority	-0.76%	2.09%	4.12%	1.05%	1.09%	3.96%
	Male	-0.18%	1.10%	3.55%	0.68%	0.53%	2.77%
	Female	-1.44%	1.35%	0.71%	-0.24%	0.53%	0.87%
Principal Job is non-S&E (OCCUP_4_I)	ALL	0.25%	-0.42%	-1.07%	-0.04%	-0.80%	-1.51%
	Bachelor	0.41%	-0.35%	-0.90%	0.01%	-0.78%	-1.58%
	Master	-0.13%	0.31%	-0.82%	-0.58%	-1.21%	-1.12%
	White	-0.04%	-1.42%	-1.49%	-0.26%	-1.47%	-1.82%
	Asian	0.94%	1.14%	-1.68%	0.55%	0.11%	-3.15%
	Minority	0.93%	0.89%	-0.07%	0.48%	0.64%	-0.22%
	Male	-0.80%	-0.73%	-2.55%	-0.79%	-1.55%	-2.56%
	Female	0.95%	-0.35%	-0.19%	0.43%	-0.34%	-0.81%

Section on Survey Research Methods

Table 1 (continued)

Items	Domain	Final weight			Adjusted weight		
		285 schools	270 schools	255 schools	285 schools	270 schools	255 schools
Working for pay or profit (WRKG_I)	ALL	0.06%	0.03%	-0.04%	0.04%	0.02%	0.12%
	Bachelor	0.05%	-0.09%	-0.28%	0.07%	-0.02%	0.07%
	Master	0.03%	0.40%	0.72%	-0.04%	0.19%	0.34%
	White	0.01%	-0.07%	-0.31%	-0.08%	-0.10%	-0.16%
	Asian	0.14%	-0.01%	0.11%	0.20%	0.11%	0.39%
	Minority	0.24%	0.32%	0.30%	0.27%	0.32%	0.26%
	Male	-0.03%	-0.04%	0.21%	-0.14%	-0.05%	0.58%
	Female	0.12%	0.10%	-0.23%	0.19%	0.09%	-0.25%
Average annual job salary (SALARY)	ALL	0.42%	0.20%	0.13%	0.33%	0.24%	0.07%
	Bachelor	0.12%	-0.24%	-0.54%	0.23%	0.15%	0.02%
	Master	0.87%	0.75%	0.93%	0.62%	0.50%	0.18%
	White	0.68%	0.93%	0.25%	0.60%	0.94%	0.13%
	Asian	-0.13%	-0.98%	0.16%	-0.30%	-0.99%	0.38%
	Minority	0.25%	-0.15%	0.14%	0.36%	-0.06%	-0.01%
	Male	0.64%	0.61%	0.82%	0.41%	0.55%	0.37%
	Female	0.19%	-0.02%	-0.50%	0.29%	-0.01%	-0.30%