# A Response Propensity Modeling Navigator for Paradata

Asaph Young Chun, Mike Kwanisai
NORC at the University of Chicago, 4350 East-West Highway, Bethesda, MD 20814

**Abstract.** The purpose of this paper is to identify factors affecting nonresponse of 12th graders in the National Assessment of Educational Progress (NAEP), by using social isolation as a theoretical navigator. In this paper, we also evaluate the statistical impact of nonresponse bias on estimates of educational performance in NAEP by taking advantage of response propensity models built on a social isolation framework. We use the 2000 NAEP science survey data and its contact history paradata, both of which are linked to the school administrative data from over 20,000 seniors in the 2000 High School Transcript Study (HSTS) whose sampling frame is identical to NAEP. We apply the final robust response propensity model to reweight NAEP estimates with additional covariates extracted from the HSTS administrative data. We evaluate the re-weighted Science performance estimates by comparing with those obtained using the current approach of NAEP nonresponse adjustment which relies on a few sampling frame variables just from NAEP data. Findings support recent research showing minimal effects on nonresponse bias of low response rates. We introduce the concept of "pandata," the data linked among multiple sources including administrative data and paradata, used for improving nonresponse adjustment methods to correct for potential nonresponse bias in survey research.

**Key Words:** Paradata, nonresponse, propensity model, complex surveys, pandata, NAEP, HSTS

## 1. Introduction

Low participation rate of students in national survey assessments increases the potential for nonparticipation bias, a product of nonparticipation rate and the difference of characteristics between participating and nonparticipating students, and thus tends to lower data credibility (e.g., Smith, 1983). Nonparticipation (or nonresponse) bias has become more important in the National Assessment of Educational Progress (NAEP), where participation rates of the 12th graders' assessment have notably declined in recent years. NAEP has been a continuing and nationally representative measure of student achievement in various subjects since 1969 in the United States. The Nation's Report Card, a major education program to document and release NAEP findings, informs the public about the academic achievement of elementary and secondary students in the U.S. The student participation rates in NAEP over the last two decades at the 12th grade have been approximately 10% to 35% lower than rates at grades 4 and 8. In 2005, the participation rate for the $12^{th}$ grade NAEP dropped to 56%, a decline of 10 percentage points from 1990. The further decline of participation rate among $12^{th}$ graders may seriously affect validity of NAEP data and in turn affect its value among key stakeholders including education policy makers. Statistical Standards[1] enforced by National Center for Education Statistics (NCES) recommends, "In cases where prior experience suggests the potential for an *overall* unit nonresponse of less than 50 percent, the decision to

[1] See U.S. Department of Education Institute of Education Sciences (2003). NCES Statistical Standards, Standard 2-2-5, page 40.

proceed with data collection must be made in consultation with the Associate Commissioner, Chief Statistician, and Commissioner."

The purpose of this paper is to investigate student- and school-factors affecting nonparticipation of 12th graders in NAEP by applying social isolation theory as guidelines and using a measurement and analysis model of nonresponse developed by Groves and Couper (1998). The research is also designed to evaluate the statistical impact of nonparticipation bias on estimates of educational performance in NAEP, using an approach used by Abraham, Maitland and Bianchi (2006). As Groves and Couper (2006) suggested, keen attention in this research is applied to investigating how strongly correlated the NAEP survey variables of interest are with (non)participation propensity, the likelihood of (non)participation. Such a research attention is justified by recent studies that demonstrated little empirical support to associate nonresponse rates with nonresponse bias (Merkele and Edelman 2002; Groves, 2006). The empirical findings might have *practical* implications about measures of interventions to adjust for nonparticipation bias and reduce nonparticipation itself in NAEP, by disclosing potential sources of nonparticipation.

Section 2 addresses a theoretical question by using a social isolation construct to explain student nonparticipation in NAEP. Nonparticipation in this research is used interchangeably with nonresponse, a term more frequently used in survey research literature. Section 3 turns to explaining data including key variables and their relevance to social isolation construct used for the research. Section 3 is where one can envision the analytical value of merging NAEP student data with the school administrative data from the High School Transcript Studies (HSTS). Because the transcripts for the 2000 HSTS are collected from *all* students in the same NAEP sample of schools regardless of individual student's participation status in NAEP, the data merged between NAEP and HSTS include key correlates of nonresponse and makes robust assessment of nonresponse bias possible. In Section 4, we present findings focused on multivariate analysis, given limited space in this paper. In Section 5, we conclude the research by elaborating implications for understanding individual and contextual factors affecting nonparticipation in NAEP, and unravel statistical impacts of nonresponse bias on NAEP estimates of educational performance.

## 2. A Modeling Navigator: Theoretical Framework

We argue, according to social isolation theory of nonparticipation (e.g., Goyder, 1987; Groves, 1989, Groves and Couper, 1998), that students perceiving or experiencing "social isolation" (e.g., those feeling not supported in family, disengaged/not motivated in classrooms, or feeling insecure/unsafe in schools) are less likely to participate in an education survey assessment, "a temporary social event" where students are assessed about knowledge gained from established social institutions. For example, a student with little motivation in classrooms is more likely to skip a class. If students with less motivation or poor performance in classroom are also less likely to participate in a NAEP assessment, student achievement in NAEP is likely to be overestimated. A student feeling insecure at schools troubled with gang activities is more likely to refuse participating in an assessment at school. At issue is how strongly correlated the assessment variables of interest are with nonparticipation propensity, the likelihood of nonparticipation in NAEP. We attempt to ground most key variables of interest in social isolation theory, as will be shown in the following sections.

Studies suggest that correlates of social isolation include demographic background factors, personal characteristics, and societal factors (Hortulanus, Machielse and Meeuwesen, 2006). Populations that are found to have high likelihoods of becoming social isolates include: the elderly, the sick and those with disability, people with lower incomes, lower educational levels, lower SES levels, and singles (e.g., Hess and Warning, 1978; Fisher, 1982). Personality characteristics that lead to becoming socially isolated include shyness, introversion, lack of social skills and the unwillingness to take social risks (Peplau and Perlman, 1982). Societal factors often associated with high social isolation include low participation in employment, club life, religious organizations, cultural activities, and volunteer work (House et al., 1982).

In order to test social isolation hypotheses in assessment survey context, we use strategies that provide us with data on key characteristics of respondents and non-respondents in NAEP by using the 2000 High School Transcript Study (HSTS) where characteristics of both participants and nonparticipants in the 2000 NAEP are contained. Because the transcripts for HSTS are collected from *all* students in the same sample of schools in which the NAEP 12th grade assessments are given, all students in NAEP assessment *including nonparticipating students* can be linked to the HSTS sample where characteristics of nonparticipants in NAEP can be studied along with that of participants from the social isolation perspective.

Participation in an assessment is an inherently tentative social process affected by personal and social factors. Thus we expect that full understanding of the process of assessment participation requires insight into key levels of influences simultaneously. We begin by exploring student- and school-level correlates of nonparticipation in NAEP by exploring variables that are justified by the construct of social isolation and evidenced in the literature. Next, we analyze the effects of key variables (i.e., student-level correlates and school-level correlates) on nonparticipation to evaluate their impacts in comparison with the current practice of NAEP merely involving some variables from the sampling frame. Finally, we model them simultaneously across levels to understand the impact of a complete set of factors on nonparticipation in NAEP.

When we turn to assessing the impact of nonparticipation bias on NAEP estimates, we use the final multivariate model of nonparticipation propensity to adjust survey weights in order to account for differences in the probability of participation associated with student- and school-level correlates, which are grounded in social isolation construct. We apply an approach Abraham, Maitland and Bianchi (2006) used for nonresponse bias analysis, so we evaluate NAEP estimates calculated using weights that incorporate our own nonresponse adjustment based on a multivariate propensity model, in comparison with NAEP estimates calculated using NAEP final weights with a nonresponse adjustment.

### *2.1* Nonresponse Bias

Best practices in surveys have been to reduce nonresponse rate in order to reduce nonresponse bias without paying due attention to the second essential component of nonresponse bias, the extent to which nonrespondents are different from respondents on statistics of interest. A traditional notion of linking high nonresponse rate to high response bias, however, has been recently challenged by several studies (Curtin, Presser, and Singer 2000; Keeter et al., 2000; Merkele and Edelman 2002) that individually demonstrated no strong relationship between nonresponse rates and nonrespnse bias. Groves (2006) further demonstrated by meta-analyzing 235 estimates from 30 studies that there is little empirical support to tie nonresponse rates to nonresponse bias. He persuasively showed that the central question is rather to investigate how strongly correlated the survey variable of interest is with response propensity, the likelihood of responding. With this perspective, the bias of the respondent mean approximates:

$B(Y_r) = Cov(Y_r, r)/R$

Where

$B(Y_r)$ = Bias of respondent mean;
$Y_r$ = Respondent population mean
$r$ = Response propensity
$R$ = Mean propensity in the target population

or

Bias (Respondent Mean) = (Covariance between survey variable, y, and response propensity, r) / (Mean propensity, R, in the target population)

Groves et al. (2006) empirically discovered that the common influences on response propensity and the survey variable of interests are reactions to the survey sponsor, interest in the survey topic, and the use of incentives. Abraham, Helms, and Presser (forthcoming) demonstrated how the strong association between the causes of volunteering and the causes of survey participation was likely to overestimate hours of volunteering in the American Time Use Survey, thus showing the significant effect of the covariance term. Further in a meta-analysis of 959 estimates from 59 studies designed to estimate the magnitude of nonresponse bias, Groves and Peytcheva (2008) concluded that high response rates are not necessarily likely to reduce the risks of bias when the cause of participation is highly correlated with the survey variables. They strongly recommended exploring how each survey variable relates to causes of survey participation in order to predict what survey estimates are most susceptible to nonresponse bias.

### 2.2 Assessing Nonresponse Bias in NAEP

Nonparticipation in the National Assessment of Educational Progress is generally the consequence of: 1) refusal by a sample student to complete the assessment, 2) failure of the sample student to be present on the day of the assessment session (absence), or 3) other reasons including the sample student's incapability to take assessment due to disability or limited English proficiency. According to the NAEP disposition guidelines Assessment Administrators use on the day of NAEP assessment, there are over 80 disposition codes of participation outcomes. In NAEP, being assessed refers to those assessed in original or makeup session with usable data. Refusal occurs when 12$^{th}$ grader or their parents (on behalf of their children) refuse to participate in the assessment. 12th graders' absence in NAEP assessment happens for various reasons: temporary (less than two weeks) or long-term illness or disability, in-school suspension, and scheduling conflicts with a sporting event usually by athletics. Other reasons of nonparticipation, according to NAEP disposition codes, are usually tied to ineligibility such as withdrawal from school or disability.

In accordance with NCES Standards 4-4-1 and 4-4-2, NAEP carries out the nonresponse bias analysis, when response rates fail to meet the required NCES standard of 85%, by using base weights for each survey stage. The existing nonresponse bias method in NAEP relies on a few school-level variables in NAEP such as type of reporting group (public vs. private school), school location (urbanicity), census region, and school size measured by student enrollment. The student-level variables selected for nonresponse bias are usually restricted to gender, age, race/ethnicity, and proxy measure of socio-economic status measured by student's eligibility for the national school lunch program. The NAEP method for assessing nonresponse bias minimally satisfies statistical standards of the National Center for Education Statistics (2003) as follows:

"*Any survey stage of data collection with a unit response rate less than 85 percent must be evaluated for the potential magnitude of nonresponse bias before the data or any analysis using the data may be released. Estimates of survey characteristics for nonrespondents and respondents are required to assess the potential nonresponse bias. The level of effort required is guided by the magnitude of the nonresponse.*"

There have been two types of nonresponse bias analysis conducted by NAEP: 1) comparison of respondents and nonrespondents across subgroups available from the sample frame, and 2) limited multivariate modeling to compare the proportional distribution of characteristics of respondents and nonrespondents to determine if nonresponse bias exists and, if so, to estimate the magnitude of the bias. The former approach is constrained by limited utility and number of frame variables which are not necessarily related to response propensity as well as variables of interest in NAEP. Asserting no evidence of nonresponse bias on the basis of similar distribution by subgroups is misleading. When this method finds certain variables associated with response, findings are reported without evaluating the direct impacts on NAEP estimates of potential nonresponse bias. The latter approach, while designed to identify the characteristics of samples least likely to respond, is limited by the extent to which predictors of interest exist only within NAEP sampling frame. For example, response outcome was modeled, in multivariate analysis, as a function of NAEP reporting group, type of school location, census region and size of school, which are all available from the sampling frame.

There have been no data available for evaluating the direct effect on NAEP achievement estimates of nonresponse bias. Nonresponse bias analysis reports prepared by NAEP have not conjectured as to the likely magnitude of any nonresponse bias in the NAEP student achievement results. Technical comments have been extremely limited in the widely used Nation's Report Cards on the perceived degree of success that has been attained in controlling NAEP nonresponse bias through the use of nonresponse adjustments. It is an untenable assumption that the sampling frame-based variables currently selected for assessing NAEP nonresponse bias are the only potential common causes affecting response propensity and NAEP statistics of interest.

## 3. Data

### 3.1 Data Sources: NAEP Survey and HSTS School Administrative Data

The two sets of data obtained from the National Center for Education Statistics are used. The first set was data from 2000 NAEP survey assessment of 12th graders and survey of their teachers and principals along with the contact history paradata among the NAEP student sample. The second set was school administrative data from 12th graders in the 2000 High School Transcript Study (HSTS). The HSTS administrative data came from all students in the same NAEP sample of schools from schools that agreed to also participate in the HSTS study, regardless of individual student participation status in NAEP. Since the HSTS data came from all students, the joint NAEP-HSTS data could be used for analysis of correlates, nonresponse and robust assessment of nonresponse bias. The joint NAEP-HSTS data was primarily used for analysis with respect to science which were assessed at $12^{th}$ graders in 2000.

The initial sample size is 23,522 students who were included in the 2000 HSTS. The NAEP-HSTS joint sample is 20,549 after dropping 1,512 students not linked to NAEP as well as ineligibles. The eligible sample of 20,549 used for this study consists of the following: 15,220 students who participated, 3,320 students who were absent, and 2,009 students who refused or whose parents refused participating in NAEP assessment on behalf of their children. Thus the weighted participation rate only at student level for the NAEP-HSTS joint sample is 75.1 percent. We remind that the HSTS student sample is obtained from NAEP schools that agreed to cooperate. If the school-level response rate is accounted for, the overall school and student combined response rate for the 2000 NAEP-linked HSTS sample would be comparable to or somewhat lower than the overall response rates of 55-60 percent in 2000 in different assessment subjects.

### 3.2 The NAEP-HSTS Joint Data

The NAEP data include score scales estimated for groups of students. The score scales are created using Item Response Theory (Lord, 1980). The NAEP score scales summarize student performance for the collection of assessment items representing the academic content specified in the NAEP frameworks specific to assessment subject. The parameters describing the item response characteristics are estimated from the score scales (Mislevy and Bock, 1982; Muraki and Bock, 1997). NAEP scores should not be compared across subjects or grades because NAEP scales are developed independently for each subject and grade.

The HSTS focuses on high school graduates' course-taking patterns, courses taken and the grades received whilst NAEP measures educational achievement in various subjects. In this paper we use the joint NAEP-HSTS data which includes variables for assessing student course-taking patterns and educational achievement. A total of 287 NAEP participating schools were in the HSTS study with 20,549 students.

In this paper we use NAEP outcome variables related to social isolation theory: 1) assessed, 2) absenteeism and 3) student and parental refusal on behalf of their children, by using potential predictors which include (a) student correlates, (b) school correlates, and (c) social psychological school climate variables, as detailed below. NAEP outcome variables are from the contact history *paradata* of over 80 official disposition codes of NAEP assessment which were classified into these major categories of participation outcome in close consultation with NCES and the NAEP participation guidelines.

Student-level correlates used were race/ethnicity, student eligibility for national school lunch program, taking advanced mathematics or science courses, GPA, Carnegie credits, standardized credits across schools, and other individual variables related to nonparticipation or student's academic performance as evidenced in literature. School-level correlates used were school urbanicity (urban, suburban, and rural), school type (public vs. private), and school enrollment size. School-wide *social psychological* correlates of nonparticipation included perception of problem activities at school, teacher absenteeism and parental support of student, which were extracted from teacher and principle surveys linked to NAEP student samples.

## 4. Results

Our analysis begins with bivariate analysis to understand the extent to which each social isolation variable is associated with nonparticipation in NAEP. We then explore the extent to which a set of variables of social isolation is likely to affect participation in NAEP in order to identify a multivariate model that is robust enough to predict participation outcomes in NAEP. Lastly we evaluate the impact on NAEP estimates of alternative nonresponse adjustment weighting that is developed from the final nonparticipation propensity model we find to be most fitting to the NAEP data. We discuss the findings, in this paper, by focusing on the final multivariate models.

.

### 4.1 Marginal Effects on NAEP Participation

In Table 1 below, we present the marginal probability effects we have generated from the multivariate logistic regressions with NAEP participation outcomes as dependent variables. Changes in predicted rates associated with having versus not having the indicated characteristics are evaluated at the overall rate for the full NAEP-HSTS sample, based on the final full logistic models of response propensity. Estimates in the 4$^{th}$ column are implied probability of contact and cooperation. Bold-faced estimates are significant at $p < .05$. For example, the figure shown in the "Low # CC (16-23)" row of the "Assessed" column in Table 1 indicates that, evaluated at the mean probability of participation (being assessed), having earned only 16-23 Carnegie credits lowers the probability of participation by an estimated 4.5 percentage points. This estimate in bold is statistically significant. This estimate, which is derived from the multivariate logistic regression with NAEP participation outcome as a dependent variable, is quite close to the implied probability of contact and cooperation, negative 4.83.

The most striking result to emerge from the data is that social isolation variables like academic indicators of Carnegie credit and GPA and school culture measures both significantly impact participation rate (being assessed) by 2 to 5 percentage points. Interestingly, school size and type, and school-level information incomplete affect the probability to be assessed by up to 20 percentage points. Other significant variables include race/ethnicity (Hispanics have higher response rate) and school urbanicity (students attending rural schools have higher response rates). These differences tend to be more affected by differences in cooperation rates, which is the similar pattern observed among 12$^{th}$ graders who are more troubled with teacher absenteeism, and lack of parental support of student achievement.

**Table 1. Marginal Effects on NAEP Participation (Being Assessed), Contact, Cooperation conditional on contact, and Comparison to Implied Probability: 2000 HSTS-NAEP, Grade 12**

| Predictor | Assessed | Contact | Cooperation, Conditional on contact | Implied Probability of Contact and Cooperation |
|---|---|---|---|---|
| (Mean of Probability) | 74.08 | 83.85 | 87.61 | 73.46 |
| Female | -1.03 | **-1.57** | 0.38 | -1.06 |
| Race/ethnicity (ref=white) | | | | |
| Black | 0.81 | **1.64** | -0.83 | 0.73 |
| Hispanic | **7.52** | **2.11** | **6.05** | 7.05 |
| Others | **8.41** | **4.15** | **5.29** | 8.29 |
| National school lunch program (ref=Ineligible) | | | | |
| Eligible for school lunch | **3.16** | 0.76 | **2.91** | 3.13 |
| Unknown | **-6.36** | **-7.59** | -0.28 | -6.86 |
| Private school | **20.10** | **11.48** | **11.50** | 21.02 |
| Census region (ref = NE) | | | | |
| Midwest | -2.33 | **-2.84** | -0.25 | -2.69 |
| South | **3.91** | **1.85** | **2.89** | 4.10 |
| West | **-4.82** | -0.12 | **-5.92** | -5.06 |
| Took advanced courses in Math or Science | **3.32** | 0.17 | **3.41** | 3.01 |
| Carnegie credits (ref = 24-28) | | | | |
| Low # CC (16-23) | **-4.50** | -0.55 | **-5.22** | -4.83 |
| High # CC (>=29) | **2.75** | **1.69** | 1.46 | 2.73 |
| No CC records | **-12.77** | **-6.91** | **-8.38** | -12.50 |
| GPA (2< ref <= 3) | | | | |
| Low GPA < =2.0 | **-4.91** | **-4.18** | -2.08 | -5.32 |
| High GPA > 3.01 | **2.42** | **3.61** | -1.09 | 2.21 |
| GPA not reported | 5.66 | **8.56** | -4.21 | 3.61 |
| Urbanicity of school location (ref = urban) | | | | |
| Suburban | 2.98 | **1.05** | 2.44 | 2.99 |
| Rural | 12.61 | **6.63** | 7.52 | 12.62 |
| School enrollment (ref = large enrollment > 900) | | | | |
| Enrollment < = 500 | 7.50 | **3.59** | 5.39 | 7.86 |
| Enrollment (501-900) | 13.11 | **8.93** | 5.87 | 13.27 |
| More problem with gang activities | 1.26 | **1.29** | 0.45 | 1.51 |
| More problem with teacher absenteeism | -5.60 | **-2.27** | -4.59 | -5.74 |
| Less parental support of student achievement | -3.55 | **-0.17** | -4.30 | -3.74 |
| School-level information incomplete | -14.15 | **-9.40** | -7.17 | -13.58 |

Note: N is 20,549. Bold-faced estimates are significant at $p < .05$. Changes in predicted rates associated with having versus not having the indicated characteristics are evaluated at the overall rate for the full NAEP-HSTS sample, based on the final logistic models of response propensity.

### 4.2 Effect on NAEP Estimates of Alternative Nonresponse Weighting Adjustments

We expect that alternative NAEP estimates derived from logistic regression models of the response propensity are in general likely to be lower than official NAEP estimates. As presented so far, we observe that students performing better, as measured by Carnegie credits or GPA, are found to be more likely to be participating in NAEP beyond and above what a number of key correlates of participation at student and school levels can account for. These correlates of proxy measure of social isolation we have conceptualized include the following: race/ethnicity, eligibility for school lunch (proxy measure of SES), school size/location/type, school-level information completeness, school characteristics as measured by school culture related to teacher absenteeism, parental support of student achievement, and problem with gang activities. We have carefully incorporated these factors into the alternative student nonresponse weight we have developed by applying logistic regression.

We also expect that alternative gap scores we re-estimate by key background variables such as race/ethnicity and school type, where we observe evidence of nonresponse bias so far, are likely to be wider. It is due to the pattern of participation in NAEP such that better performing students are found to be more likely to participate and poor performing students are less likely to participate, beyond and above what can be explained by a set of student factors (race/ethnicity, gender, eligibility for national school lunch) as well as school-level variables (school climate measures, school size, type, urbanicity, and location). The participation propensity scores we have incorporated into the alternative student nonresponse adjustment weighting factor reflect such a pattern of participation. Thus we expect the NAEP achievement gap is likely to be wider in alternative weighting method, especially where background measures are found to be significant predictors of participation of $12^{th}$ graders in NAEP.

We calculate the estimated participation propensity for each NAEP participant based on the final full logistic regression coefficients. We compute the student nonresponse adjustment weight by taking the inverse of the estimated response propensity for each participating $12^{th}$ grader in NAEP. Using the propensity-score-based weight adjustment, we recalculate NAEP estimates of scale score in the 2000 NAEP Science. We perform analysis with WesVar to properly account for the complex multi-stage clustered NAEP sample design and to re-estimate NAEP scale scores with alternative nonresponse adjustment. We also adjust a set of replicate weights by a factor of alternative nonresponse weighting to produce proper standard errors of re-estimated NAEP scale scores.

**Table 2. Effects of Weights on Estimates of Mean NAEP Scale Scores in Science, 2000 HSTS-linked NAEP at Grade 12**

| | Science (0-300 scale) | | | |
|---|---|---|---|---|
| | NAEP Final Weight | | Own Final Weight with Alternative Nonresponse Adjustment | |
| | **Score** | **SE** | **Score** | **SE** |
| Overall Mean | 146.6 | 1.0 | 145.4 | 1.0 |
| Male | 147.6 | 1.3 | 146.4 | 1.3 |
| Female | 145.6 | 1.1 | 144.5 | 1.1 |
| (Male-Female) | 2.0 | 1.3 | 1.9 | 1.4 |
| White | 152.9 | 1.1 | 152.1 | 1.1 |
| Black | 121.5 | 1.8 | 121.0 | 1.7 |
| Hispanics | 129.9 | 2.0 | 129.4 | 2.1 |
| Others | 150.5 | 3.7 | 148.8 | 3.1 |
| (White - Black) | 31.5* | 2.0 | 31.1* | 1.9 |
| (White - Hispanics) | 23.0* | 1.9 | 22.7* | 2.2 |
| (White - Others) | 2.4 | 3.7 | 3.3 | 3.0 |
| Northeast | 149.4 | 2.8 | 148.8 | 2.7 |
| Midwest | 150.0 | 1.7 | 149.3 | 1.8 |
| South | 142.4 | 1.3 | 141.6 | 1.2 |
| West | 147.4 | 2.9 | 143.8 | 2.7 |
| (NE - Midwest) | -0.6 | 3.3 | -0.5 | 3.3 |
| (NE - South) | 7.0* | 3.1 | 7.1* | 2.9 |
| (NE - West) | 2.1 | 4.0 | 4.9 | 3.8 |
| Public | 145.1 | 1.0 | 143.8 | 1.0 |
| Private | 163.5 | 1.5 | 163.3 | 1.6 |
| (Private - Public) | 18.4* | 1.9 | 19.5* | 1.9 |

**NOTE**: A single asterisk indicates the gap score is statistically significant at $p < .05$.

Table 2 above summarizes re-estimated NAEP Science scale scores in comparison with the official NAEP estimates produced, using the current NAEP weights developed for the 2000 NAEP Science. Estimates in the table include NAEP scale scores overall and by key background variables, and achievement gap by key variables such as gender, race/ethnicity, and school type. Standard errors of estimates are included in the second column under each set of data. NAEP scale score results are a numeric summary of what students know and can do in a particular subject. Science estimates are on a scale of 0 to 300. Achievement gap describes student achievement in terms of the gap, for example, between Black and White students, between Hispanic and White students, and between male and female students. Evaluating achievement gap by key background variables is the essence of the “No Child Left Behind” mandates. Key education policies at the federal level are guided by their impacts on reducing such an achievement gap.

The most notable pattern in this table appears to be about how closely NAEP scale scores lie between estimation methods using NAEP final weight and our own alternative weight. Gap scores are found to be little affected except for the census region of West. Reweighting lowers science scores for both male and female students, thus not affecting the gender gap much. The science scores by race/ethnicity are generally lower than official estimates of NAEP. Thus the achievement gap between White and other races is not affected. The only exception is the achievement gap widened between White and others that include Asian-Pacific American and American Indian students. Reweighting appeared to widen the achievement gap between students in private and public schools, with an increase of over 1-point. Reweighting widened the regional gap of science scores, in particular between schools in the Northeast and the West, getting more than twice wider (2.1 points vs. 4.9 points). The reader is cautioned that given the size of associated standard errors, the observed change may be small.

## 5. Discussion and Conclusions

We began this research motivated by the relatively low response rate of NAEP at $12^{th}$ grade (i.e., about 10% to 35% lower than rates at grades 4 and 8). We were concerned about the potential for nonresponse bias in NAEP estimates due to the difference between participants and nonparticipants in NAEP or the extent of covariance between NAEP variables of interest and response propensity, as Groves and Couper (2006) theorized. We explored from this research empirical implications in response propensity models of identifying student- and school-level factors affecting nonparticipation of $12^{th}$ graders in NAEP. We examined NAEP estimates for $12^{th}$ graders by applying the approach used by Abraham, Maitland, and Bianchi (2006) to evaluate the impact of nonresponse bias on NAEP estimates.

The analysis provides evidence on the origins and the implications NAEP nonparticipation associated with this broad context of nonresponse research we began.
First, we have investigated nonresponse bias, using a concept of social isolation (or social integration) to identify a set of variables applied to developing response propensity models. We have analyzed to the NAEP 2000 data a social isolation construct which Groves and Couper (1998) applied initially in household surveys. It can be seen as a social integration approach to building nonresponse models proposed by Lepkowski and Couper (2002) and by Abraham, Maitland, and Bianchi (2006). The social isolation framework has been applied to investigate how a set of factors determining 12th graders' participation in NAEP might be useful to evaluate their effects on sequential process of participation involving contactability and cooperation conditional on contact. The contactability model takes into account absence (i.e., noncontact); the cooperation model, refusal either by students or by parents on behalf of their children. We observed that the contribution of absence to NAEP nonparticipation is about 50% higher than for refusal by students and their parents. The utility of linking NAEP survey and paradata to the HSTS school administrative data is demonstrated by testing the social isolation hypotheses and designing approaches to improving nonresponse bias analysis and in turn designing nonresponse adjustment. We are inclined to call multiple data linked together with a modeling navigator (e.g., social isolation, social integration) as "pandata," perhaps a new concept that is subject to further investigation. We demonstrated the merits of pandata by linking three sets of data (i.e., NAEP survey, NAEP contact history paradata, and HSTS school administrative data) for a particular statistical purpose of nonresponse adjustment.

It should be noted that this research, constrained by lack of direct measures of social isolation, could include such a social psychological measure of social isolation, using scales of shyness, introversion, and lack of social skills. It is also desirable to measure school-level factors of social isolation/integration by tapping students' involvement in study groups, after-school activities, religious organizations, and volunteer activities in order to associate the scope of these voluntary activities with participation in NAEP.

Second, we find evidence of significant relationships between participation and a number of student- and school-level variables, but no evidence that reweighting the data in the fashion as suggested by alternative response propensity models has affected the NAEP estimates. We have found evidence confirming the covariance between NAEP variables of interest and response propensity. Namely we observed a significant relationship of response propensity with measures of academic achievement (e.g., Carnegie credit and GPA) and contextual measures of school culture (e.g., perception of problem with teacher absenteeism and parental support of student achievement), respectively. In the science NAEP, we observe student achievement as measured by GPA plays an essential role in predicting participation rates in the context of controlling for a set of student- and school-level variables as used for science NAEP.

However, when the response propensity models derived from multivariate logistic regressions are applied to re-estimating NAEP scale scores, there is no evidence that reweighing the data has a significant or meaningful effect on the NAEP estimates in science. That is not a ground to rule out nonresponse bias in NAEP estimates, since other subject-specific student- or school-level variables could account for the differences between participants and nonparticipants. Reweighting with our own alternative nonresponse adjustment has lowered the science mean estimates by approximately 1-point on a scale of 0-300. When comparing NAEP estimates calculated from the official NAEP weight and our own alternative weight, the achievement gap in NAEP science appears to be pretty close to each other by gender and race/ethnicity. The science achievement gap gets a little wider when comparing the private-public achievement gap, and it gets apparently wider when evaluating regional differences, in particular between schools in the Northeast and West.

This research extends the findings by Curtin, Presser and Singer (2000) in demonstrating minimal damage of nonresponse bias. A traditional notion of linking high nonresponse rate to high response bias has been also challenged by Keeter et al. (2000) and Merkle and Edelman (2002) who showed no strong relationship between nonresponse rates and nonrespnse bias. Groves (2006) further demonstrated this by meta-analyzing 235 estimates from 30 studies that there is little empirical support to associate nonresponse rates to nonresponse bias. Findings from the current research with NAEP data strengthen such an argument. NAEP scores in 2000 science reweighted with our response propensity model would not affect most of statistical inferences made about achievement gap by key variables in the year 2000, as the net effects on NAEP scores appear not to be large. Previous NAEP publications in science indicate that even one-point of scale score can on occasion make a difference especially when it is about the achievement gap by such key variables as gender, race/ethnicity, and eligibility for national school lunch program (a proxy measure of poverty). Thus it is premature to judge how meaningful or meaningless apparently small gap scores would be.

Third, it might be useful to develop in the future a nonparticipation index, an indicator of participation difficulty. This indicator may be constructed on the basis of a response propensity model of student- and school-level variables. Such a nonparticipation index may be linked specifically to the origins of nonparticipation -- student refusal, parental refusal, and student absence -- so that corresponding conversion strategies can be effectively developed in the NAEP field of data collection. NCES recently reported that the response rate of NAEP at grade 12 has been increased in the 2007 Writing Assessment, speculating it was perhaps due to design changes, best practice guidelines that recently began (e.g., offering more make-up sessions of NAEP assessment at school), or demographic shifts in the student population. However, it is not empirically possible to confirm which of design changes or best practice has contributed to increasing the response rate at grade 12. No experimental studies have been carried out to test the impact of individual NAEP features on increasing response rate. Twelfth graders' absence in NAEP assessment happens for various reasons. Empirical findings support that Black 12th graders attending large public schools in urban areas are more likely to be absent, compared to peers in other race/ethnicity groups. If students in this school setting are more encouraged for participation by additional make-up sessions, it should reduce a potential bias due to noncontact in particular.

## Acknowledgements

We would like thank the National Center for Education Statistics for supporting research presented in this paper by providing access to the restricted-use NAEP and HSTS data. Our thanks are extended particularly to Dr. Peggy Carr, Associate Commissioner, and Dr. Andy Kolstad, Senior Technical Advisor of NCES.

## References

Abraham, K.G., Helms, S.E., and Presser, S. (Forthcoming). How social processes distort measurement: the impact of survey nonresponse on estimates of volunteer work in the United States. *American Journal of Sociology*.

Abraham, K.G., Maitland, A., and Bianchi, S.M. (2006). Non-response in the American Time Use Survey: Who is Missing from the Data and How Much Does It Matter? Public Opinion Quarterly, 70, 676-703.

Curtin, R., Presser, S. and Singer, E. (2000). "The Effects of Response Rate Changes on the Index of Consumer Sentiment." Public Opinion Quarterly, 64:413-428.

Goyder, J. (1987). The Silent Minority: Nonrespondents on Sample Surveys. Boulder, CO: Westview Press

Groves, R.M. (1989). Survey errors and survey costs. New York: John Wiley and Sons, Inc.

Groves, R. M. (2006), Nonresponse rates and nonresponse bias in household surveys, *Public Opinion Quarterly,* 70, 646-675.

Groves, R.M., and Couper, M.P. (1998). *Nonresponse in household interview surveys.* New York: John Wiley and Sons, Inc.

Groves, R.M., Couper, M.P., Presser, S, Singer, E., Tourangeau, R., Acosta, G.P., and Nelson, L. (2006). Experiments in Producing Nonresponse Bias, *Public Opinion Quarterly*, 70, 720-736.

Groves, R. M. and Peytcheva, E. (2008), The impact of nonresponse rates on nonresponse bias, *Public Opinion Quarterly*, 72, 167-189.

Hortulanus, R, Machielse, A, and Meeuwesen, L. (2006). Social isolation in modern society. NY: Routledge.

House, J.S., Robbins, C. and Metzner, H.L (1982). The association of social relationships and activities with morality: Prospective evidence from the Tecumseh community health study. American Journal of Epidemiology, 116. 123-140.

Keeter, S., Kohut, A., Miller, A., Groves, R., and Presser, S. (2000). Consequences of reducing non-response in a large national telephone survey. Public Opinion Quarterly, 64(2), 125-48.

Lepkowski, J.M. and Couper, M. (2002). "Nonresponse in the second wave of longitudinal household surveys," in Groves, R., Dillman, D., Eltinge J., and Little R. (eds), Survey Nonresponse, 259-274, New York: Wiley.

Lord, F.M. (1980). Applications of Item Response Theory to Practical Testing Problems. Hillsdale, NJ: Lawrence Erlbaum Associates.

Merkle, D. and Edelman, M. (2002). "Nonresponse in exit polls: a comprehensive analysis," in Groves, R., Dillman, D., Eltinge J., and Little R. (eds), Survey Nonresponse, pp 243-258, New York: Wiley.

Mcevoy, A., and Welker, R. (2000). Antisocial behavior, academic failure, and school climate: a critical review. Journal of Emotional and Behavioral Disorders

Mislevy, R.J., and Bock, R.D. (1982). BILOG: Item Analysis and Test Scoring with Binary Logistic Models [Computer program]. Chicago: Scientific Software.

Muraki, E., and Bock, R.D. (1997). PARSCALE: IRT Item Analysis and Test Scoring for Rating-Scale Data. Chicago, IL: Scientific Software International.

National Center for Education Statistics (2003). NCES Statistical Standards, Institute of Education, U.S. Department of Education, NCES 2003-601. Washington, DC.

Neuman, S., Hagedorn, T., Celano, D., and Daly, P. (1995). Toward a collaborative approach to parent involvement in early education: a study of teenage mothers in an African community. American Educational Research Journal, 32 (4), 801-827.

Peplau, L. A. and Perlman, D. (1982). Loneliness: a sourcebook of current theory, research and therapy. NY: John Wiley.

Presser, S. (1981). "A comparison of Harris sample composition with census, IRS, and NORC," In E. Martin, D. McDuffee, and S. Presser (eds.), Source book of Harris National Surveys: Questions 1963-76. Chapel Hill, North Carolina; Institute for Research in Social Science.

Roey, S., Caldwell, N., Rust, K., Hicks, L., Lee, J., Perkins, R., Blumstein, E., and Brown, J. (2005). The 2000 High School Transcript Study User's Guide and Technical Report (NCES 2005–483). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Skinner, E.A., and Belmont, M.J. (1993). "Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year." *Journal of Educational Psychology, 85*(4).

Smith, T. (1983). "The Hidden 25 percent: an analysis of nonresponse on the 1980 General Social Survey," Public Opinion Quarterly, 47, 1.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics (2000). The High School Transcript Study. Washington, D.C.