

**Texas Higher Education Opportunity  
Project (THEOP) Sophomore Wave 2  
“Stayer/Leaver” Survey**

**Methodology Report**

February 17, 2005

# TABLE OF CONTENTS

---

Introduction	4
Survey Background	4
Survey Objectives	4
Study Procedures	5
Survey Universe	5
Sampling Frame	5
Sample Design and Selection	5
Race/Ethnicity Disproportionate Sampling	6
Sampling Strategy	7
The Survey Instrument	8
Field Protocol	8
The Field Period	8
Results	9
Wave 2 Respondents Compared to Total Wave 2 Sample and Non-Respondents	11
Data Processing	12
Analytic Weights	13
Questionnaire	15

# LIST OF TABLES AND FIGURES

---

Table 1: Expected Distribution of Sophomore Race/Ethnicity when sampling proportionate to race/ethnicity and under a recommended alternative	6
Table 2: Racial and Ethnic Origin of Respondents – Unweighted and Weighted	7
Table 3: Sample	7
Table 4: Stayers/Leavers Screened	9
Table 5: Stayers/Leavers Completed Survey	9
Table 6: Students That Ever Dropped Out of School	9
Table 7: Call Outcomes	10
Table 8: Response Rate by Phase	10
Table 9: Baseline Racial/Ethnic Origin of Wave 2 Respondents and Non-Respondents	11
Figure 1: Baseline Gender of Wave 2 Respondents and Non-Respondents	11
Figure 2: Baseline Born in United States for Wave 2 Respondents and Non-Respondents	12
Figure 3: Baseline Language Other Than English Spoken at Home for Wave 2 Respondents and Non-Respondents	12

# INTRODUCTION

---

This report documents the design and implementation of the Texas Higher Education Opportunity Project 2004 Sophomore Wave 2 “Stayer/Leaver” Survey. The survey is the first follow-up with Baseline Sophomores who completed questionnaires in school the spring of 2002.

## SURVEY BACKGROUND

The Texas Higher Education Opportunity Project conducted baseline surveys with sophomores in the state of Texas in the spring of 2002. The survey population consisted of all sophomores attending public high school in the state. Students attending charter schools, those in special education classes (or schools) and those attending schools with fewer than 10 sophomores were excluded. The baseline interviews were mainly collected in class using a self-administered scannable survey booklet, although a small portion of students were mailed the survey booklet and returned it by mail. A total of 19,969 surveys were completed with high school students that were sophomores in the 2001 to 2002 school year. The project will continue to refer to this cohort as “sophomores” as they progress through the waves of the longitudinal study even though the term no longer describes the students’ level in school.

## SURVEY OBJECTIVES

The Sophomore Wave 2 “Stayer Leaver” Study is the first follow-up with a subsample of baseline sophomores. Most of the respondents were in their senior year of high school at the time of the interview. The focus of the survey is on the student’s activities during the senior year and plans after high school. An important component of this study was to partition the sophomore cohort into Stayers and Leavers.

**Stayers:** Stayers represent those students who have attended the same high school from the baseline survey in 2002 to the wave 2 survey in 2004. Analysis of students who stayed at the same high school will determine whether students’ knowledge of the Top 10 Percent law increased and whether they changed their college aspirations as they progressed through school.

**Leavers:** Leavers are those students that have changed schools or dropped out (and did not return to the same high school) between the baseline survey and the wave 2 survey. Analysis of the leaver students will determine whether, how many, and which students deliberately changed schools in order to qualify for the benefits of the Top 10 Percent law.

Students that ever dropped out of school, regardless of whether they returned to school or not, were asked a series of questions that explored reasons for dropping out and activities during their time away from school. Students that dropped out, but then returned to the same high school are defined as Stayers. Those that dropped out and did not return to school, or attended a different school, are defined as Leavers.

# STUDY PROCEDURES

---

## SURVEY UNIVERSE

The survey universe for the sophomore wave 2 study was defined as all students who participated in the baseline survey, signed the assent form, and could be identified by name. The vast majority of sophomore baseline participants provided at least some information on the contact information sheet included in the baseline survey. For those participants who signed the assent form but failed to provide any contact information, we assigned the home address from school directories to the extent possible and also conducted research, mainly via the web, for contact information. Baseline respondents who were anonymous (gave no identifying information) or who failed to sign the written assent form were not included for selection.

## SAMPLING FRAME

The sampling frame was the survey database of 19,969 students that completed the 2002 Baseline Sophomore Survey rather than the contact information file because the principal stratification variable (race/ethnicity) was obtained from responses to survey questions, and IRB protocols prohibit the association of respondent answers with respondents' identities. Students were eligible for inclusion in the wave 2 study if they participated in the baseline survey, signed the assent form and could be identified by name. An "eligibility flag," containing an indicator of whether or not the subject was eligible to be included, was merged from the contact information file onto a subset of the survey data (containing stratification variables, school code and student ID number). The sample was drawn from this extracted and processed data set, with the student ID serving as the link between the sample and the corresponding contact information.

## SAMPLE DESIGN AND SELECTION

The survey objectives of the first follow-up study called for a total of 3,000 completed interviews with a subsample of sophomores who participated in the spring 2002 Texas Higher Education Opportunity Project Baseline Survey. Specifically, the precision requirements call for 2,000 completed surveys with stayers and 1,000 completed surveys with leavers to accommodate the need to analyze the stayer and leaver subgroups separately as well as to contrast the two groups. In addition, comparative analysis was desired for the following demographic subgroups:

- Non-Hispanic Whites
- African Americans
- Hispanics
- Asians.

**Disproportionate sampling of leavers.** A key sampling issue was the need to sample the leaver and stayer groups differentially to effect the desired sample size targets. Random sampling of baseline sophomores was expected to yield leavers in the range 600 to 900 – short of the 1,000 target specification. Thus, oversampling was deemed necessary with the allowance that the oversampling plan should be flexible to accommodate the measure of uncertainty associated with the prevalence of leavers.

## RACE/ETHNICITY DISPROPORTIONATE SAMPLING

Apart from the leaver/stayer domains, race/ethnicity represent important analytic subgroups to be considered. The expected distribution of race/ethnicity on a proportionate sample of 3,000 subjects from the sophomore baseline survey data was calculated and the results are shown in Column A of Table 1 (using unweighted data). A recommendation was made, and followed, to increase the representation of Blacks, Hispanics and Asians (irrespective of leaver/stayer status) so that race/ethnic subgroup specific analyses can be conducted. Column A shows that less than 100 Asians and 400 Blacks would be available for analyses under proportionate sampling by Race/Ethnicity. An increase in the representation of these groups (as well as Hispanics) was suggested to ensure a nominal ability to analyze these important subgroups. The recommendation shown under Column B was followed. It should be noted that three quarters of the “All Else” row are subjects who failed to provide race/ethnicity on their baseline questionnaire. The majority of these respondents did report race/ethnicity during the wave 2 survey.

**TABLE 1: EXPECTED DISTRIBUTION OF SOPHOMORE RACE/ETHNICITY WHEN SAMPLING PROPORTIONATE TO RACE/ETHNICITY AND UNDER A RECOMMENDED ALTERNATIVE**

	<b>A</b>	<b>B</b>
<b>Race/Ethnicity:*</b>	<b>Proportionate**</b>	<b>Recommended</b>
<b>White</b>	<b>1190</b>	<b>1057</b>
<b>Black</b>	<b>353</b>	<b>400</b>
<b>Hispanic</b>	<b>980</b>	<b>1000</b>
<b>Asian</b>	<b>91</b>	<b>200</b>
<b>All else</b>	<b>386</b>	<b>343</b>
<b>Total</b>	<b>3000</b>	<b>3000</b>

\* Distribution is based on original (unedited) responses to Q46 of the Baseline Survey

\*\* Assumes the sample is drawn proportionately across race/ethnicity (using actual counts of available cases).

Additional Caveats/Assumptions: The expected distributions did not take into account: (1) student written assent; and (2) availability of contact information (i.e., the back sheet of the survey instrument). Some sample loss will occur as a result, since such cases are not eligible to be put into the field. Also, these distributions implicitly assume that leaver/stayer status is not associated with race/ethnicity.

Table 2 shows the racial and ethnic origin of wave 2 **respondents** both unweighted and weighted by the analytic weight (based on TEA enrollment figures) that allows inference to the total population.

**TABLE 2: RACIAL AND ETHNIC ORIGIN OF RESPONDENTS – UNWEIGHTED AND WEIGHTED**

	Unweighted Frequency	Unweighted Percent	Weighted (to TEA population) Frequency	Weighted (to TEA population) Percent
White	1,137	37 %	104,556	39 %
African American or Black	431	14 %	45,980	17 %
Mexican, Mexican American or Chicano	1,088	35 %	83,431	31 %
Other Hispanic	171	6 %	14,516	6 %
Asian or Pacific Islander	175	6 %	9,010	3 %
Native American	13	<1 %	1,224	1 %
Something Else	72	2 %	6,868	3 %
Refused	5	<1 %	664	<1 %
<b>TOTAL</b>	<b>3,092</b>	<b>100 %</b>	<b>266,249</b>	<b>100 %</b>

## SAMPLING STRATEGY

A two-phase sampling approach was employed in the execution of the survey to facilitate the field operations for stayer and leaver data collection. The approach relied on the efficient use of sample replicates. A total of 11,295 subjects were released into the field.

**Replication** - The oversampling plan involved partitioning the sophomore sample into 27 replicates, each representing a random subsample of the whole. Replicates were then released into the field for dialing. The early “unrestricted” replicates allowed both stayers and leavers to be eligible for interviewing. Later “restricted” replicates only allowed leavers and drop-out stayers to be eligible for interviewing. In these latter replicates, screening questions quickly established the eligibility of the subject and terminated contact with ineligible stayers.

This sampling strategy effectively oversampled leavers, and allowed substantial flexibility in fine-tuning the stayer sample to ensure attainment of the targeted 2,000 completed interviews. The use of replicates also helped boost response rates by allowing replicates to be more fully worked before releasing additional sample replicates.

Table 3 shows that 3,410 pieces of sample were used in Phase I interviewing, while 7,885 pieces were released in Phase II.

**TABLE 3: SAMPLE**

	Frequency	Percent
Phase I – All are eligible	3,410	30%
Phase II – Only Leavers and Stayer Dropouts are eligible	7,885	70%
<b>Total</b>	<b>11,295</b>	<b>100%</b>

## PARENTAL CONSENT

Unlike the Senior Follow-up survey of 2003, some of the sophomore cohort were minors, and most were still attending high school at the time of the interview. Thus, parental consent was required for minors prior to assent by the student. The birth year was present for over 90% of the sample, therefore the need for parental consent was calculated by the CATI (Computer Aided Telephone Interviewing) program based on the date each sample piece was called since students were usually celebrating their 18<sup>th</sup> birthday during the interviewing period. Parental consent was required on all cases where the birth year or other components of the birth date were not present. We spoke to the parent or guardian for approximately one quarter of the sample and received approval for all but three cases (i.e. three cases ended with a final disposition of Guardian Refusal).

## THE SURVEY INSTRUMENT

The survey instrument was a computer-assisted-telephone interviewing (CATI) questionnaire approximately 20 minutes in length. An introductory paragraph informed respondents that the survey was being conducted on behalf of Princeton University and was part of a study about educational opportunity in the state of Texas. Respondents were reminded their participation was totally voluntary. The questionnaire (see Appendix) covered the following:

1. Screening to determine Stayer/Leaver Status
2. Dropout Interview
3. Course Taking/Grades
4. Test Scores and Guidance
5. Knowledge and Perception of College Admissions
6. Future Plans
7. Self Esteem, Agency and Control
8. Demographic Information

We also gathered contact information verification for those that received an incentive.

## FIELD PROTOCOL

The field protocol for the Sophomore Wave 2 Stayer/Leaver Survey included the following:

- 1) Issue an advance letter to students with "mailable" addresses
- 2) The advance letter contained a \$2 incentive as well as a post card for updating contact information
- 3) Telephone students with usable telephone numbers and conduct CATI instrument
- 4) Conduct locating efforts for students lacking a valid telephone number using:
  - Contact information provided by student
  - Student directories obtained from schools
  - Free web based person finders
  - Subscription based person finders on the web
- 5) Offer incentives to increase participation rates in the later months of the field period

## THE FIELD PERIOD

Following this field protocol, 9,884 pieces of sample were loaded into the CATI system in February 2004. The Phase I sample of 3,410 cases was dialed on immediately, while the Phase II sample was held in reserve. Dialing on Phase II sample began in mid-April after analysis of Phase I suggested that the 2,000 stayer completes would be achieved from the replicates released in Phase I. Dialing continued on Phase I sample that had not been brought to a final disposition, such as "completed survey," throughout the entire

field period. In mid-May the determination was made to add additional Phase II sample (1,411 pieces brought the total sample to 11,295) to reach the goal of 1,000 leavers. The overall field period for telephone interviewing was February 26 through October 6, 2004.

## RESULTS

Nearly 75 percent (8416/11,295) of the total sample was screened for this survey (see Table 4). In the Phase II portion of the study, parents were allowed to complete the first screening question if the student was not available.

**TABLE 4: STAYERS/LEAVERS SCREENED**

	Frequency
Stayers	7,413
Leavers	1,003
<b>Total</b>	<b>8,416</b>

The goal of 2,000 stayers was met, and even exceeded, while the goal of 1,000 leavers was nearly met (achieving 97% of the target; see Table 5). For the cases that were screened as leavers but did not complete the survey, an average of 32 call attempts were made ranging from 10 to 87 attempts.

**TABLE 5: STAYERS/LEAVERS COMPLETED SURVEY**

	Frequency
Stayers	2,122
Leavers	970
<b>Total</b>	<b>3,092</b>

Table 6 shows that 272 students that had ever dropped out of school were interviewed. Of those, fifteen were classified as “Stayer Dropouts” meaning they dropped out of school, but then returned to the same school. “Leaver Dropouts” (n=257) are those students that dropped out of school and either never returned, or went back to a different school than the school they were at for the baseline survey.

**TABLE 6: STUDENTS THAT EVER DROPPED OUT OF SCHOOL**

	Frequency
Stayer Dropouts	15
Leaver Dropouts	257
<b>Total</b>	<b>272</b>

Incentives were used to increase participation rates at the tail end of the field period. When we first started using incentives, \$10 was offered. This increased to \$20 within a few weeks. In the final days of data collection, the incentive was increased to \$30. Eighty-two respondents received a \$10 incentive, 544 respondents received \$20, and four respondents received a \$30 incentive.

Table 7 shows the call outcomes for the total sample and Table 8 shows outcomes for sample based on Phase I/Phase II assignment.

**TABLE 7: CALL OUTCOMES**

CALL OUTCOMES	Frequency	Percent
Completed Full Survey	3092	27.4%
Completed Screening for Stayers	5165	45.7%
Guardian Did Not Consent to Survey	3	0%
Final Refusal from Respondent	78	0.7%
Respondent Deceased	6	0.1%
Special Education Student	1	0%
Exchange Student	1	0%
Did Not Complete Screener or Survey (Approximately 1/5 of these Students were Located, 4/5 Not Located)	2,949	26.1%
<b>TOTAL SAMPLE</b>	<b>11,295</b>	<b>100%</b>

Response rates are calculated independently for each phase of interviewing (see Table 8).<sup>1</sup> For Phase I, all respondents were eligible to participate in the survey so the response rate is the same as the interview rate or 70.5 percent. For Phase II the response rate is calculated by multiplying the screening response rate by the interview response rate; this yields an overall Phase II response rate of 72.3 percent.

The final response rate for the overall study can be calculated by taking a weighted average of the Phase I and II response rates (which reflect the amount of sample in each phase). By multiplying the Phase I response rate times 0.3 (since 30 percent of the total sample was assigned to Phase I) and adding the result to the Phase II response rate times 0.7 (70 percent of the total sample was assigned to Phase II), we produce an overall response rate of 72 percent for the total study.

**TABLE 8: RESPONSE RATE BY PHASE**

	Frequency	Percent
<b>TOTAL PHASE I – All are Eligible</b>	<b>3410</b>	<b>100%</b>
Unscreened	869	25.5%
Screened	2541	74.5%
Completed Survey	2403	70.5%
<b>RESPONSE RATE (interview rate since all are eligible)</b>		<b>70.5%</b>
<b>PHASE II – Only Leavers and Dropouts are Eligible</b>	<b>7885</b>	<b>100%</b>
Not Screened – Unknown if Eligible	2010	25.5%
Screened	5875	74.5%
Known Ineligible	5165	
Known Eligible	710	
Completed Survey	689	97.0%
<b>RESPONSE RATE (screening rate x interview rate)</b>		<b>72.3%</b>
<b>FINAL RESPONSE RATE ((0.30*.705) + (0.70*.723))</b>		<b>71.8%</b>

<sup>1</sup> Students that were not able to participate in the study (n=8) were left in the response rate calculations rather than deleting them from the base because the numbers were so small as to be negligible.

## WAVE 2 RESPONDENTS COMPARED TO TOTAL WAVE 2 SAMPLE AND NON-RESPONDENTS

The **baseline** characteristics of survey respondents (those that completed the full interview or the screener for stayers) were compared to the **baseline** characteristics of non-respondents for a few key variables. Baseline data (collected in the Spring 2002 baseline survey) was used for this analysis because there is no wave 2 data for the non-respondents.

Figure 1 shows that the gender split of wave 2 respondents is very close to the total of the wave 2 sophomore sample. Non-respondents were evenly split between male and female as reported on the baseline survey with slightly higher missing data for this question than the total wave 2 sample.

FIGURE 1: BASELINE GENDER OF WAVE 2 RESPONDENTS AND NON-RESPONDENTS

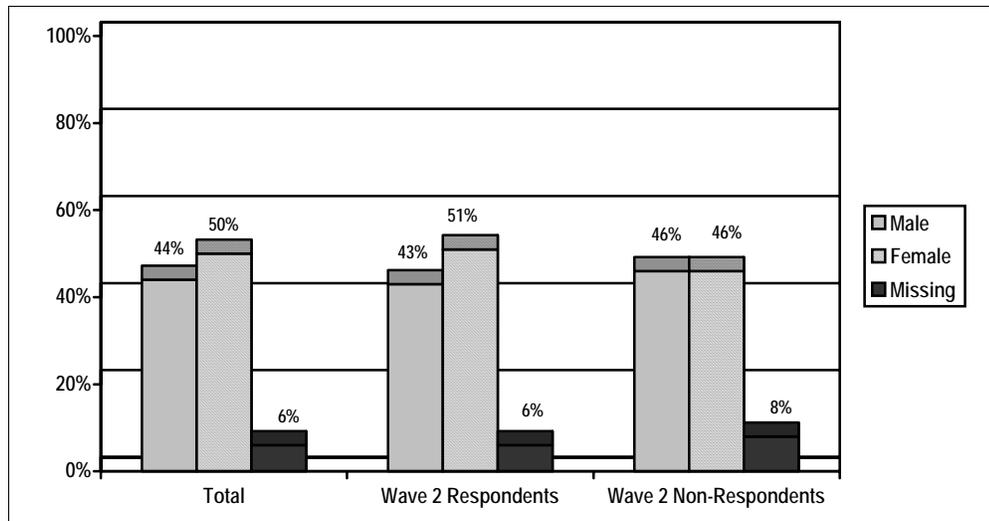


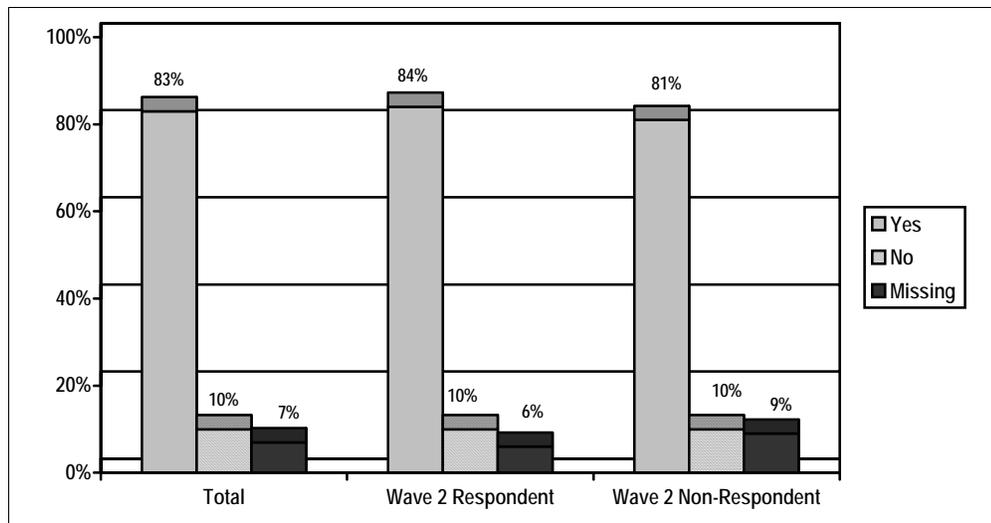
Table 9 contains the racial and ethnic origin of wave 2 sophomore sample as reported in the baseline survey, not the wave 2 survey. This data is not weighted. Wave 2 respondents are very similar to the total sample while non-respondents show slight differences for Whites and African Americans but overall the percentages are close.

TABLE 9: BASELINE RACIAL/ETHNIC ORIGIN OF WAVE 2 RESPONDENTS AND NON-RESPONDENTS

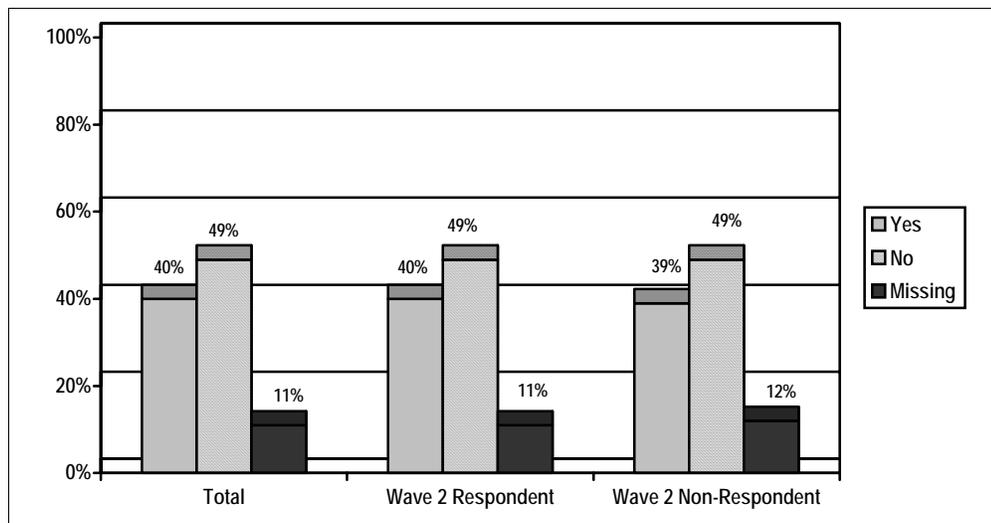
	Total Percentage	Wave 2 Respondent Percentage	Wave 2 Non-Respondent Percent
White	37%	38%	32%
African American	13%	12%	16%
Mexican/Mexican American/Chicano	30%	30%	29%
Other Hispanic	5%	5%	5%
Asian or Pacific Islander	5%	5%	5%
Native American	1%	1%	1%
Other	3%	3%	4%
Missing	7%	6%	9%
<b>Total</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>
<b>N</b>	<b>11295</b>	<b>8257</b>	<b>3038</b>

There was some concern that students who were not reachable for the wave 2 survey may have returned to the country of origin, in particular Mexico. Figure 2 and Figure 3 allay this concern. Both respondents and non-respondents are almost identical to the total wave 2 sample on the baseline characteristics of birth in the United States and speaking a language other than English at home.

**FIGURE 2: BASELINE BORN IN UNITED STATES FOR WAVE 2 RESPONDENTS AND NON-RESPONDENTS**



**FIGURE 3: BASELINE LANGUAGE OTHER THAN ENGLISH SPOKEN AT HOME FOR WAVE 2 RESPONDENTS AND NON-RESPONDENTS**



## DATA PROCESSING

Survey data were processed to ensure that all data items were valid, or reflected missing or inapplicable codes, as appropriate.

## ANALYTIC WEIGHTS

Analytic weights were developed for the survey data. The weights are comprised of two components:

- a sampling weight -- the selection probability of being sampled into the Sophomore Follow-up; and
- a post-stratification adjustment that corrects for student level and school level nonresponse.

The sampling weight is simply the reciprocal of the probability of selection. For the sophomore follow-up, the selection probability of a student is simply the product of the initial Baseline school selection probability ( $\pi$ ) and the probability of selection of the student into the Follow-up. Thus, the probability of a student ( $\lambda$ ) in a high school is simply equal to:

$$\lambda_{ik} = \pi_i \times \omega_{i,j}$$

where  $\pi$  denotes the baseline school selection probability and  $\omega$  is the probability that student ( $j$ ) in school ( $i$ ) is selected into the 2004 Sophomore follow-up sample. The sampling weight is simply the inverse of the  $\lambda$  for each student ( $k$ ):

$$w_k = 1 / \lambda_k$$

The post-stratification adjustment aligns the school level student totals to published TEA enrollment figures within the original Top 10 sampling strata. For each of the sampling strata ( $h$ ), known totals of sophomore students were calculated using the TEA master frame used to draw the original baseline sample.

The post-stratification adjustment,  $A(h)$ , was developed by taking ratio of *published* population totals  $T(h)$  for each stratum to the *weighted survey totals*  $S(h)$  for each corresponding stratum  $h$ :

$$A(h) = T(h)/S(h)$$

where

$$T(h) = \sum_i t_{ih}$$

and where  $i$  denotes summation across all schools in the TEA school file that belong to Stratum  $h$ . Also,

$$S(h) = \sum_i w_i \delta_{ih}$$

where  $\delta$  denotes an indicator variable that is equal to 1 if a student is in a school belonging to Stratum  $h$  and equals 0 otherwise. The  $i$  denotes summation across all students in the sophomore survey file whose schools belong to Stratum  $h$ .

The *final analytic weight*  $\theta$  is the product of the selection probability and the post-stratification weights. For each student  $i$ ,

The adjustments incorporate student level nonresponse. Moreover, the results ensure that the weighted

$$\theta_i = w_i \times A_i$$

data will align with the published totals across sampling strata.

**Sampling errors.** Because the sample design employed a two-stage unequal probability sample design, the estimates of statistical precision provided by statistical packages that presume a simple random sample cannot be employed. The complex nature of the sample must be taken into account.

# QUESTIONNAIRE

---

The CATI questionnaire is posted on the THEOP website and can be accessed at:  
<http://theop.princeton.edu/surveys.html>