Texas Higher Education Opportunity Project (THEOP) Senior Wave 2 Survey

Methodology Report

January 23, 2004

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

This report documents the design and implementation of the Texas Higher Education Opportunity Project Senior Wave 2 Survey. The survey is the first follow-up with Baseline Seniors.

SURVEY BACKGROUND

The Texas Higher Education Opportunity Project conducted baseline surveys with seniors in the state of Texas in the spring of 2002. The survey population consisted of all seniors 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 seniors 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 13,803 surveys were completed with high school students that were seniors in the 2001 to 2002 school year. The project will continue to refer to this cohort as "seniors" 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 Wave 2 Senior Study is the first follow-up with a subsample of baseline seniors. This phase tracks the evolution of student decision-making about college attendance among those who decide to attend college (full or part time) immediately after high school graduation as well as those who decide to attend college one or more years after graduation. The survey also covers post high school activities including military enlistment, employment, civic activities, high school experiences, life events, self-esteem and current living status. The following demographic subgroups will be used for comparative analyses:

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

Additionally, separate analyses are desired for:

- Students attending college or technical school
- Students not attending college one year after attending high school.

STUDY PROCEDURES

SURVEY UNIVERSE

The survey universe for the Senior Wave 2 Study was defined as all students who participated in the baseline survey, signed the consent form, and could be identified by name. The vast majority of senior 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 rather than the contact information file because the principal stratification variables are obtained from responses to survey questions, and IRB protocols prohibit the association of respondent answers with respondents' identities. 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 6,000 completed interviews with a subsample of seniors who participated in the spring 2002 THEOP Baseline Survey. In addition to conducting statewide analyses, the analytic goals of the study are driven by a need to conduct comparative analyses among the following demographic subgroups: non-Hispanic Whites, African Americans, Hispanics, and Asians as well as separate analyses for students who are attending college or technical school and those who are not attending college one year after attending high school.

A key sampling issue was the need to oversample certain subgroups to meet the survey objectives (specifically, that of conducting subgroup analysis for the specified domains -- racial-ethnic groups; college and non-college kids). Therefore the strategy of oversampling race-ethnic subgroups was selected. A census was taken of all available, eligible African Americans and Asians and students intending to enter the military. For the remainder of sample, subsamples were drawn by a proportionate sample design. Table 1 presents the sampling targets by Follow-up sampling strata (each cell of the table defines a subsampling stratum).

TABLE 1: Target Sample Sizes for the Follow-up of Baseline Seniors by Subsampling Strata*

Race/ethnicity:	White	Black	Hispanic	Asian	All else	Total
College	1568	***802****	994	***358****	312	*4035*
Tech school	54	***52****	89	***9****	24	*229*
Work	148	***87****	183	***17****	46	*481*
Military	***78****	***55****	***95****	***7****	***24****	***258****
All else	170	***104****	261	***24****	438	*997*
Total	*2019*	***1100****	*1622*	***414****	*845*	*6000*

*** Cells in asterisks indicate certainty subsampling strata; that is, all baseline respondents were sampled into the follow-up in these cells.**

Table 2 shows the Wave 2 **respondents** by the **Baseline** subsampling strata. Note, data used for the subsampling strata is from the Baseline survey, not the Wave 2 survey, and as with all follow-up surveys, some respondents may (and did) answer the race/ethnicity question differently. Wave 2 actual respondents very closely mirror the expected results seen in Table 1.

TABLE 2: Wave 2 Respondents by Baseline Subsampling Strata

Race/ethnicity:	White	Black	Hispanic	Asian	All else	Total
College	1629	744	1054	331	307	*4065*
Tech school	44	37	91	6	16	*194*
Work	125	64	178	15	36	*418*
Military	126	38	152	5	30	*351*
All else	123	95	242	20	328	*808*
Total	*2047*	*978*	*1717*	*377*	*717*	*5836*

Table 3 presents the baseline characteristics of the Wave 2 **non-respondents**.

TABLE 3: Wave 2 Non-Respondents by Baseline Subsampling Strata

Race/ethnicity:	White	Black	Hispanic	Asian	All else	Total
College	565	334	339	157	107	*1502*
Tech school	31	32	30	5	16	*114*
Work	80	24	77	5	20	*206*
Military	70	35	86	3	23	*217*
All else	104	58	115	12	182	*471*
Total	*850*	*483*	*647*	*182*	*3448*	*2510*

Turning from Baseline characteristics to Wave 2 reported data, table 4 shows the racial and ethnic origin of Wave 2 **respondents** both unweighted and weighted.

TABLE 4: RACIAL AND ETHNIC ORIGIN OF RESPONDENTS – UNWEIGHTED AND WEIGHTED

	Unweighted Frequency	Unweighted Percent	Weighted Frequency	Weighted Percent
White	2,229	38.2	102,347	48.8
African American or Black	1,028	17.6	20,804	9.9
Mexican, Mexican American or Chicano	1,630	27.9	58,640	28.0
Other Hispanic	335	5.7	11,001	5.3
Asian or Pacific Islander	390	6.7	7,865	3.8
Native American	18	0.3	684	0.3
Something Else	195	3.3	7,909	3.8
Refused	11	0.2	267	0.1
TOTAL	5,836	100	209,518	100

THE SURVEY INSTRUMENTS

The main survey instrument was a computer-assisted-telephone interviewing (CATI) questionnaire approximately 25 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 A) covered the following:

1. **Post High School Activity Status**

2. **Activities**
   - 2.1 Military
   - 2.2 Employment
     - 2.2.1 Currently working
     - 2.2.2 Not working, but worked since senior year
     - 2.2.3 Looking for work
   - 2.3 Schooling
     - 2.3.1 Since September 2002
     - 2.3.2 Between June and August 2002

3. **College Sequence**
   - 3.1 Applying and Admission
   - 3.2 School Expenses & Finances
   - 3.3 School Experiences
   - 3.4 Future Plans and Educational Expectations

4. **Civic Activities**
5. **High School Experiences**
6. **Negative Life Events**

7. **Self Esteem, Agency and Control**
8. **Current Living Status and Demographics**

CONTACT INFORMATION VERIFICATION

In addition a military proxy questionnaire was used (see Appendix B) so that a proxy could answer a small number of questions on behalf of the respondent who is away on duty in the military and unreachable during the field period of the study. The proxy questionnaire consisted of 17 questions. These cases will be periodically called during 2004 to attempt to complete the full interview with the respondent.

FIELD PROTOCOL

The field protocol for the Wave 2 Survey of Baseline Seniors involved the following process:

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) Special strategies
   - Assign cases to teams of interviewers to work exclusively – this allows the interviewers to become more familiar with each case and build greater rapport with screeners, gatekeepers, information providers, and respondents.
   - Send emails to email addresses collected in the baseline study asking the student to contact us via email or our 800 in-bound telephone number
   - Assign laptops to selected interviewers to conduct interviews outside of normal calling hours as desired by the respondent
6) Conduct interviews with proxies for respondents unreachable due to service in the military
7) Offer incentives to increase participation rates in the later months of the field period

THE FIELD PERIOD

Following this field protocol, an initial sample of 8,302 cases was selected and loaded into the CATI system in April 2003. Additional research located telephone numbers for 44 more cases that were added to the CATI system in July for a total sample size of 8,346. The 44 additional cases did not receive an advance letter.

Telephone interviewing began April 3 and continued until December 9, 2003. Military proxy responses were collected from September 18 to November 26, 2003, and four cases were

converted from military proxy completes into full survey completed interviews between December 19 and December 27, 2003.

RESULTS

Over 131,000 calls were made to complete 5,836 interviews. An average of fifteen call attempts were made per piece of sample. For completed surveys, the average was 12 call attempts while for those cases that never completed a survey the average is over 20 call attempts, ranging from 1 attempt to 99 attempts. The average length of completed full interviews was approximately 25 minutes (ranging from 9 minutes to 64 minutes).

Incentives were used to increase participation rates at the tail end of the field period. For military proxy surveys $5 was sent with a thank-you letter to generate goodwill among proxies that should lead to an increased respondent participation rate in the full survey. For full surveys $10 was sent in a thank-you letter for surveys completed from September 15 to October 16. From October 17 to December 9, a $20 incentive was used. The $20 incentive will also be offered to military proxy cases as we attempt to complete full surveys with respondents throughout 2004.

A total of 5,836 interviews were completed with 5,741 completed as full interviews with the respondent and 95 interviews completed with proxies for respondents that are unreachable due to military service (see Table 5). Only 159 respondents (2%) refused to participate in the study and 5 cases were ineligible. Of the remaining sample, roughly two-thirds could not be located to complete the survey while one third were believed to be 'located' but did not complete the survey during the field period.

TABLE 5: CALL OUTCOMES

Call Outcome	Frequency	Percent
Completed Full Survey	5741	69
Completed Proxy Survey	95	1
TOTAL COMPLETES	5836	70
REFUSED TO PARTICIPATE	159	2
Deceased	4	0
Special Education Student	1	0
TOTAL INELIGIBLE	5	0
REMAINING SAMPLE (1/3 Located, 2/3 Not Located)	2346	28
GRAND TOTAL	8346	100

The following tables are based on unweighted data from the Senior Wave 2 Study. Respondents were split almost equally by gender with 48% male and 52% female (Table 6).

TABLE 6: GENDER

	Frequency	Percent
Male	2791	47.8
Female	3040	52.1
Not Sure	5	0.1
Total	5,836	100.0

Over 98% of respondents had graduated from high school (Table 7).

TABLE 7: GRADUATED FROM HIGH SCHOOL

	Frequency	Percent
Yes	5,745	98.4
No	91	1.6
Total	5,836	100.0

Of all respondents, including military proxy surveys, 318 (5.4%) have been sworn into the armed forces (Table 8).

TABLE 8: MILITARY SERVICE

	Frequency	Percent
Army	94	1.6
Navy	56	1.0
Marine Corps	64	1.1
Air Force	59	1.0
Coast Guard	2	.0
Army Reserves	15	.3
Navy Reserves	1	.0
Marine Corps	7	.1
Air Force Reserves	2	.0
Air National Guard	1	.0
Army National Guard	13	.2
Other	4	.1
Not in Military	5,518	94.6
Total	5,836	100

Of those respondents that reported, in the Baseline survey, their primary activity in the fall after leaving high school would be joining the military, roughly half (55%) had joined the military by the time of the Wave 2 survey (see Table 9).

TABLE 9: BASELINE PRIMARY ACTIVITY IN THE FALL BY MILITARY SERVICE

BASELINE Primary Activity in the Fall	Joined Military	Did Not Join Military	Total
College	77	3988	4065
Tech School	4	190	194
Work	11	407	418
Military	193	158	351
All Else	33	775	808
Total	318	5518	5836

Over three quarters of respondents reported working for pay since the end of their senior year in high school (Table 10). Just under half of respondents have worked in one job while a third have had more than one job (Table 11).

TABLE 10: WORKED FOR PAY SINCE SENIOR YEAR IN HIGH SCHOOL

	Frequency	Percent
Yes	4546	77.9
No	1195	20.5
Missing (Not asked of surveys completed by proxy)	95	1.6
Total	5,836	100.0

TABLE 11: NUMBER OF JOBS SINCE END OF SENIOR YEAR

	Frequency	Percent
One	2597	44.5
Two	1489	25.5
Three	347	5.9
Four	84	1.4
Five	17	0.3
Six	5	0.1
Seven or More	7	0.1
None	1195	20.5
Missing (Not asked of surveys completed by proxy)	95	1.6
Total	5,836	100.0

Three quarters of respondents said they attended some type of school since September 2002 (Table 12) and two-thirds were attending school at the time of the interview (Table 13).

TABLE 12: ATTENDED A VOCATIONAL, TECHNICAL, OR TRADE SCHOOL OR TAKEN COURSES FROM A UNIVERSITY OR COLLEGE FOR ACADEMIC CREDIT SINCE SEPTEMBER 2002

	Frequency	Percent
Yes	4328	74.2
No	1413	24.2
Missing (Not asked of surveys completed by proxy)	95	1.6
Total	5,836	100.0

TABLE 13: CURRENTLY ATTENDING SCHOOL

	Frequency	Percent
Yes	3943	67.6
No	1798	30.8
Missing (Not asked of surveys completed by proxy)	95	1.6
Total	5,838	100.0

Just under 14% of respondents said they were granted admission to a Texas college or university through the Top 10 Percent Law (Table 14).

TABLE 14: GRANTED ADMISSION TO ANY TEXAS COLLEGE OR UNIVERSITY THROUGH THE TOP 10 PERCENT LAW

	Frequency	Percent
Yes	808	13.8
No	4933	84.5
Missing (Not asked of surveys completed by proxy)	95	1.6
Total	5,836	100.0

Eleven percent of respondents agreed that the Top 10 Law was one of the reasons they decided to attend their principal institution (or apply if they were not attending school) (Table 15).

TABLE 15: TOP TEN LAW WAS A REASON FOR ATTENDING INSTITUTION

	Frequency	Percent
Yes	645	11.0
No	5096	87.3
Missing (Not asked of surveys completed by proxy)	95	1.6
Total	5,836	100.0

Roughly two thirds of respondents (63%) provided their social security number. This will be extremely useful for tracking students for additional waves of the study.

DATA PROCESSING

Survey data were processed to ensure that all data items were valid, or reflected missing or inapplicable codes, as appropriate. As expected in all surveys of this size and type, some anomalies did occur.

ANALYTIC WEIGHTS

Analytic weights were developed for the 2003 senior follow-up survey data. The weights reflect two components:

- a sampling weight
- post-stratification adjustment that corrects for student level and school level nonresponse.

The sampling weight is simply the reciprocal of the probability of selection. The selection probability of a student is equal to the product of the student's baseline selection probability and the follow-up stratum-specific subsampling probability used to define sample size targets for the follow-up sample. Letting the baseline selection probability of the student *i* in subsampling stratum *k* be π, and the stratum specific subsampling rate be defined by λ, then overall selection probability of a student (ν) is:

$$\nu_{ik} = \pi_i \times \lambda_{ik}$$

The values of the subsampling weight, $1/\lambda$, for each follow-up stratum are presented in the Table 16 below:

TABLE 16: SUBSAMPLING WEIGHTS BY FOLLOW-UP SAMPLING STRATA

Intentions:	White	Black	Hispanic	Asian	All else
College	1.827	1.0	1.827	1.0	1.827
Tech school	1.844	1.0	1.835	1.0	1.853
Work	1.668	1.0	1.670	1.0	1.409
Military	1.0	1.0	1.0	1.0	1.0
All else	1.971	1.0	1.936	1.0	1.872

Table 16 reflects the certainty selection of Blacks, Asians, and those indicating intentions to join the military after attending high school.

The sampling weight is simply the inverse of the ν for each student (*i*):

$$w_i = 1/\nu_i$$

The post-stratification adjustment aligns the school level student totals to published TEA enrollment figures within our original Baseline sampling strata. For each of the 11 sampling strata (*h*), known totals of senior and sophomore students were calculated using the TEA master frame used to draw the sample.

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

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

where

$$T(k) = \sum_i t_{ik}$$

(i denotes summation across all schools in the TEA school file that belong to PSU *k*) and

$$S(k) = \sum_i w_i \delta_{ik}$$

(δ denotes an indicator variable that is equal to 1 if a student is in a school belonging to PSU *k* and equals 0 otherwise; the i denotes summation across all students in the survey file whose schools belong to PSU *k*).

The *final analytic weight* θ is the product of the selection probability and the post-stratification weights:

For each student *i.*

$$\theta_i = w_i \times A_i$$

The post-stratification adjustments ensure that the weighted data will align with the published totals across the original baseline sampling strata.

QUESTIONNAIRES

The CATI questionnaire (including the military proxy questionnaire) is posted on the THEOP website and can be accessed at: http://theop.princeton.edu/surveys.html