High School Longitudinal Study of 2009 (HSLS:09) Panel Maintenance 2018 & 2021

Appendix C: Responsive Design Supplement

OMB# 1850-0852 v.28

National Center for Education Statistics
U.S. Department of Education

June 2018

Appendix C: Responsive Design Supplement

This appendix provides supplementary details on the development and results of the responsive design approach used in the High School Longitudinal Study of 2009 (HSLS:09) second follow-up main study. This appendix is intended to complement the material in section 4.2 which provides detailed coverage of the data collection design and responsive design strategy implemented in the second follow-up. In this appendix, the following specific sections are provided: section F.1 summarizes the second follow-up responsive design approach used; section F.2 details the development of the two responsive design models employed, the response likelihood model (F.2.1) and the bias likelihood model (F.2.2); section F.3 provides the results of the calibration sample experiments; and section F.4 reports on the effects of the responsive design approach on key survey estimates.

C.1 Second Follow-up Responsive Design

An advantage of the responsive design approach is that it allowed for periodic assessment, during data collection, of how representative the responding sample was of the total population represented in the study so that efforts and resources could be focused on encouraging participation among the cases that were most needed to achieve representativeness in the responding sample. The approach implemented in the HSLS:09 second follow-up was designed to increase the overall response rate in a cost-sensitive, cost-efficient manner and that also reduces the difference between respondents and nonrespondents among key variables, thereby more effectively reducing the potential for nonresponse bias. An uninformed approach to increase response rates may not successfully reduce nonresponse bias, even if higher response rates are achieved (Curtin, Presser, and Singer 2000; Keeter et al. 2000). Decreasing bias during the nonresponse follow-up depends on the approach selected to increase the response rate (Peytchev, Baxter, and Carley-Baxter 2009). In the current approach, nonresponding sample members who were underrepresented among the respondents were identified using a statistical model (bias likelihood model) which incorporated covariates that were deemed relevant to the reported estimates (e.g., demographic characteristics and key variables measured in prior survey administrations). Once identified, these critical nonrespondents could be targeted for tailored incentives dependent on their respective subgroup.

The second follow-up sample was divided into three subgroups of interest, based on prior experience with the cohort, so that customized interventions could be developed based on patterns of response behavior from prior data collection rounds and applied to each group independently. The subgroups consisted of the following:

1. **Subgroup A** (high school late/alternative/noncompleters [HSNC]) contained the subset of sample members who, as of the 2013 Update, had not completed high school, were still enrolled in high school, received an alternative credential, completed high school late, or experienced a dropout episode with unknown completion status.

2. **Subgroup B** (ultra-cooperative respondents [UC]) consisted of sample members who participated in the base year, first follow-up, and 2013 Update without an incentive offer. These cases were also early web respondents to the 2013 Update and on-time or early regular high school diploma completers.[1]

3. **Subgroup C** (high school completers and unknown high school completion status [HS other]) included cases that, as of the 2013 Update, were known to be on-time or early regular diploma completers (and not identified as ultra-cooperative) and cases with unknown high school completion status that were not previously identified as ever having had a dropout episode.

To determine optimal incentive amounts, a calibration subsample was selected from each of the aforementioned subgroups to begin data collection ahead of the main sample. The experimental sample was treated in advance of the remaining cases. Results from the calibration sample experiments were used to determine the incentive levels – a *baseline* incentive and two subsequent incentive increases, or *boosts* – offered to the remaining (i.e., noncalibration) sample in each of the three subgroups.

[1] In the spirit of a responsive design, the set of cases to be treated as "ultra-cooperative" was expanded for the main sample (i.e., cases not in the calibration sample) with the goal of maximizing the efficient use of project resources because response rates were reasonably high. See section 4.2.1.6 for further details and for the expanded definition. The definition provided above corresponds to that used for sample members in the calibration sample.

The data collection design for the second follow-up included a responsive design with multiple intervention phases. These phases included specific protocols for handling each of the three subgroups of sample members to reduce the potential for biased survey estimates or reduce data collection costs (Peytchev 2013). For more details on the second follow-up data collection design, see section 4.2.1.

C.2 Responsive Design Model Development

In the HSLS:09 second follow-up, two models were used to help identify, or target, cases for specific interventions. The models consisted of an estimated a priori probability of response for each member (assigned using a *response likelihood* model) and a *bias likelihood* model to identify nonrespondents in underrepresented groups. The bias likelihood model identified which cases were most needed to balance the responding sample. The response likelihood model helped to determine which cases were optimal for pursuing with targeted interventions so that project resources could be most effectively allocated.

C.2.1 Response Likelihood Model Development

The response likelihood model was developed using data from earlier rounds, and was designed to predict the a priori likelihood of a case becoming a respondent. The response likelihood model allowed the data collection team to identify cases with a low probability of responding and avoid applying relatively expensive interventions, such as field interviewing, to these cases. To make the interventions more cost efficient, the primary objective of the response likelihood model was to inform decisions about the exclusion of cases that were identified for targeting based on the bias likelihood model but which had extremely low likelihood of participation. From a model-building perspective, the objective was to maximize prediction of participation, regardless of any association between the predictor variables and the HSLS:09 survey variables.

From prior analysis in the base year, first follow-up, and 2013 Update, candidate variables known to be predictive of response

behavior (i.e., prior-round response outcomes) were considered for the response likelihood model. To determine which covariates to include in the model, stepwise logistic regression was run with the model entry criteria set to $p = .5$—meaning that any predictor variable with an initial probability value of .5 or less was included in the stepwise regressions—and model retention criteria set to $p = .1$—meaning that any variable with a probability value of .1 or less was retained in the final model. The result of this approach is the retention of a set of covariates capable of predicting a case's likelihood of becoming a respondent. Table F-1 lists all predictor variables considered for inclusion in the response likelihood model and their final inclusion disposition (i.e., which variables were retained and which were released from the final model).

Table C-1. Candidate variables for the response likelihood model and final retention status: 2016

Data source	Variable	Retention status
Sampling frame	Sex	Retained
	Race/ethnicity[1]	Retained; no significant differences in likelihood of response between White sample members and Asian sample members. All other race/ethnicity comparisons to White sample members were significant.
Base year	Response outcome	Retained
First follow-up	Response outcome	Retained
Panel maintenance updates / Other update activities	First follow-up panel maintenance response outcome	Retained
2013 Update	Response mode	Not retained
	Ever called in to the help desk	Not retained
	Ever agreed to complete web interview	Retained
	Ever refused (sample member)	Retained
	Ever refused (other contact)	Retained
	Phase targeted and incentive amounts	The following variables were retained: 1) Case offered a $40 baseline incentive (ever-dropouts) 2) Case offered the abbreviated interview 3) Case was never targeted with any incentive The incentive boost amounts and the prepaid incentive variables were not included in the final model.
	Dual language speaker	Retained
	High school diploma status	Retained
	Completed high school on time	Retained

[1] Race categories exclude persons of Hispanic ethnicity.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09), Base Year, First Follow-up, 2013 Update, and Second Follow-up.

Response likelihood model results. The odds ratio, confidence interval, and interpretation of each covariate are presented in table F-2. The odds ratios describe how much more likely a case is to be a respondent than a nonrespondent.

Table C-2. Odds ratios and confidence intervals for variables in the response likelihood model: 2016

			95% confidence interval		
Data source	**Variable**	**Odds ratio**	**Lower bound**	**Upper bound**	**Interpretation**
Sampling frame	Sex	1.17	1.069	1.280	Females were more likely to respond than males
	Race/ethnicity: Hispanic compared to White	0.74	0.645	0.854	Hispanics were less likely to respond than Whites
	Race/ethnicity: Black compared to White	0.80	0.682	0.913	Blacks were less likely to respond than Whites
	Race/ethnicity: Other compared to White	0.80	0.686	0.931	Other race/ethnicities were less likely to respond than Whites
Base year	Response outcome	1.60	1.415	1.885	Base-year respondents were more likely to respond than base year nonrespondents
First follow-up	Response outcome	3.39	3.002	3.798	First follow-up respondents were more likely to respond than first follow-up nonrespondents
Panel maintenance update	First follow-up panel maintenance response outcome	1.74	1.559	1.939	First follow-up panel maintenance respondents were more likely to respond than first follow-up panel maintenance nonrespondents
2013 Update	Ever agreed to complete the web survey	2.66	2.196	3.227	Cases that ever agreed to complete the web survey were more likely to respond than those that had not agreed
	Ever refused (sample member)	0.09	0.080	0.110	Cases that ever refused were less likely to respond than those that had not refused
	Ever refused (other contact)	0.08	0.070	0.088	Refusals by other were less likely to respond than those who never refused
	Case offered a $40 baseline incentive (ever-dropout)	1.89	1.611	2.217	Ever-dropout cases offered $40 incentive were more likely to respond than those offered other incentive amounts
	Case offered the abbreviated interview	0.04	0.037	0.050	Cases offered the abbreviated interview were less likely to respond than those not offered the abbreviated interview

See notes at end of table.

Table C-2. Odds ratios and confidence intervals for variables in the response likelihood model: 2016—Continued

Data source	Variable	Odds ratio	95% confidence interval: Lower bound	95% confidence interval: Upper bound	Interpretation
	Case was never targeted with an incentive offer	0.44	0.386	0.490	Cases never targeted were less likely to respond than those that were targeted
	Dual language status	1.47	1.275	1.689	English-only speakers were more likely to respond than those of other languages
	High school diploma status	2.18	1.601	2.971	High school diploma recipients were more likely to respond than those that had not earned a high school diploma
	Completed high school on time	3.72	2.744	5.042	On-time high school completers were more likely to respond than those who had not completed high school on time

NOTE: Race categories exclude persons of Hispanic ethnicity.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09), Base Year, First Follow-up, 2013 Update, and Second Follow-up.

Response likelihood model definition. Using the final covariates selected (primarily paradata variables), a model was developed to predict the response outcome in the 2013 Update, the last data collection round prior to the second follow-up. The response likelihood model used a logit function to generate, for each case, a continuous probability of response (bounded by 0 and 1), called a response likelihood score, in which a value of 1 indicated a case was predicted to respond and 0 indicated a case was predicted not to respond. Response likelihood values were calculated one time prior to the beginning of data collection.

We label the 2013 Update survey responses, $y_i^{2013\,\text{Update}}$, as 1 for respondents and 0 for nonrespondents and model them with $\Pr(y_i^{2013\,\text{Update}}=1)=\text{logit}^{-1}(X_i\beta)$. Input variables are modeled as independent and include sex (female), prior-round response status (e.g., base year response), and the remaining retained covariates specified in table F-2. This model, therefore takes the expanded form

196 $$\Pr(y_i^{2013\ \text{Update}} = 1) = \text{logit}^{-1}(\beta_0 + \beta_1 \bullet \text{female}_i + \beta_2 \bullet \text{base year response}_i + \ldots)$$

197 From this model, we derive predicted response likelihood scores,
198 $\hat{p}_i^{\text{response}}$, for each case, defined as

199 $$\hat{p}_i^{\text{response}} = \frac{e^{(\beta_0 + \beta_1 \bullet \text{female}_i + \beta_2 \bullet \text{base year response}_i + \ldots)}}{1 + e^{(\beta_0 + \beta_1 \bullet \text{female}_i + \beta_2 \bullet \text{base year response}_i + \ldots)}} = \Pr(y_i^{2013\ \text{Update}} = 1)$$

200 **Overall response likelihood distribution.** Across the entire second
201 follow-up fielded sample (n = 23,316)[2], the overall mean response
202 likelihood score was .80. As indicated by this mean, many sample
203 members were clustered at the upper end of the distribution. Within
204 the three subgroups of interest, subgroup A (HSNC; n = 2,545) had a
205 mean response probability of .65. As expected, these cases were
206 found to have the lowest average response likelihood value among
207 all of the subgroups. Conversely, subgroup B (UC; n = 4,144) had a
208 mean response probability of .96, indicating that these cases were
209 highly likely to be respondents per the response likelihood model.
210 Subgroup C cases (HS other; n = 16,627) had a mean response
211 probability of .78, very close to the fielded sample's overall mean.

212 As noted in section 4.2.1.2, the model-derived response likelihood
213 scores were used to assist in determining intervention resource
214 allocation only in phases 5 and 6 to avoid pursuing cases in field
215 interviewing that were unlikely to respond. Section 4.2.1.4 provides
216 further details on the use of these scores.

217 C.2.2 Bias Likelihood Model Development

218 The goal of the bias likelihood model was to identify cases most
219 likely to contribute to nonresponse bias because their characteristics
220 were underrepresented among the set of respondents. This approach
221 provided an overview of where sample underrepresentation might
222 be occurring in the respondent set. To achieve this goal, the criteria
223 for inclusion of variables in the bias likelihood model differed from
224 the criteria for inclusion in the response likelihood model.
225 Maximizing the prediction of survey participation was not the main
226 objective. In the bias likelihood model, variables of high analytic
227 value were sought for inclusion in the model. Therefore, model fit

15 [2] See section 2.4 for a description of the second follow-up sample design.

and statistical significance were not primary determining factors in deciding which variables to include in the bias likelihood model. Rather, variables were selected for inclusion in the bias likelihood model principally due to their analytic importance to the study. Conversely, variables that were highly predictive of participation but not necessarily associated with the survey variables, such as paradata on the ease of obtaining participation on the previous administration, were excluded as they could have a disproportionate influence on the predicted propensities without contributing additional information on bias in the second follow-up. Once the set of key variables was identified, stepwise logistic regression was used to help improve overall model fit. Bias likelihood model variables, and their corresponding level of data requiring imputation, are presented in table F-3. Note that many key survey variables from prior rounds contained missing values which required imputation to be included in the bias likelihood model. Further discussion of the imputation process follows in the text below.

Table C-3. Bias likelihood model variables: 2016

Data source	Variable	Percentage of cases requiring imputation
Sampling frame	Sex	No missing data; imputation not required
	Race/ethnicity[1]	No missing data; imputation not required
	School type	No missing data; imputation not required
	School locale (urbanicity)	No missing data; imputation not required
Base Year	How far in school 9th grader thinks he/she will get	12.0
	How far in school parent thinks 9th grader will go	28.4
	9th grader is taking a math course in the fall 2009 term	9.5
	9th grader is taking a science course in the fall 2009 term	9.5
	Mathematics quintile score	8.8
First follow-up	Teenagers final grade in algebra 1	14.3
	How far in school sample member thinks he/she will go	12.0
	How far in school parent thinks sample member will go	10.5
	Grade level in spring 2012 or last date of attendance	12.6
	Student dual language indicator	0.4
	Socioeconomic status composite	10.5
	Teenager has repeated a grade	10.8
	Mathematics quintile score	12.0

See notes at end of table.

Table C-3. Bias likelihood model variables: 2016—Continued

Data source	Variable	Percentage of cases requiring imputation
2013 Update and High School Transcript Collection	Teenager has high school credential	20.4
	Taking postsecondary classes as of Nov. 1, 2013	20.7
	Level of postsecondary institution as of Nov. 1, 2013	21.2
	Apprenticing as of Nov. 1, 2013	20.8
	Working for pay as of Nov. 1, 2013	20.8
	Serving in military as of Nov. 1, 2013	21.0
	Starting family/taking care of children as of Nov. 1, 2013	20.9
	Number of postsecondary institutions applied to	22.7
	Currently working for pay	21.5
	Number of high schools attended	6.0
	Attended CTE center	6.0
	English-language learner status	6.0
	GPA: overall	6.1
	GPA: English	6.1
	GPA: mathematics	6.2
	GPA: science	6.2
	Total credits earned	6.0
	Credits earned in academic courses	6.0

[1] Race categories exclude persons of Hispanic ethnicity.

NOTE: GED = general educational development; FAFSA = Free Application for Federal Student Aid; CTE = career and technical education; GPA = grade point average.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09), Base Year, First Follow-up, 2013 Update, High School Transcript Study, and Second Follow-up.

Imputation process. Assessment of balance between respondents and nonrespondents required having nonmissing data for both groups. To be used as bias likelihood model covariates, many key survey variables containing missing values required imputation. Missing data were imputed for these survey variables using stochastic imputation. Prior-round nonrespondents were included in imputation since the goal was to achieve a complete dataset for all second follow-up sample members. Specifically, a weighted sequential hot-deck (WSHD) statistical imputation procedure (Cox

1980; Iannacchione 1982), using the student bas e weight[3], was applied to the missing values for the variables. The WSHD procedure replaces missing data with valid data from a donor (i.e., item respondent) within an imputation class, or what is commonly called a donor pool. For nonrespondents with all missing survey data from a prior data collection round (i.e., prior-round nonrespondents), frame data – available for all sample members – were used to form donor pools which were used to impute missing survey data.

Imputation classes were identified using a recursive partitioning function (also known as a nonparametric classification tree, or classification and regression tree [CART], analysis) through the *tree* (Ripley 2015) package in R (R Core Team 2015). In addition to the survey items used to form imputation classes, sorting variables were used within each class to increase the chance of obtaining a close match between donor and recipient. If more than one sorting variable was chosen, a serpentine sort[4] was performed where the direction of the sort (ascending or descending) changed each time the value of a variable changed. The serpentine sort minimized the change in the respondent characteristics every time one of the variables changed its value. With recursive partitioning, the association of a set of survey items and the variable requiring imputation is statistically tested (Breiman et al. 1984). The result was a set of imputation classes formed by the partition of the survey items that are most predictive of the variable in question. The pattern of missing items within the imputation classes was expected to occur randomly, allowing for the WSHD procedure to be used (note that the WSHD procedure assumes data are missing at random within imputation classes). Input items included the sampling frame variables and survey variables imputed earlier in the ordered sequence, or those that were identified through skip patterns in the instrument, or through literature suggesting an association.

[3] The student base weight was used as it is nonmissing for all sample members. For further details on weights available in the second follow-up, including the student base weight, see chapter 6.

[4] A serpentine sort is a sorting method in which records are ordered in an alternating ascending and descending pattern, thereby causing any two consecutive records in the sorted file to have similar values for the sort variables.

Finally, the student base weight was used to ensure that the population estimates calculated post-imputation did not change significantly from the estimates calculated prior to imputation. Missing values were successfully imputed for the majority of the variables, allowing them to be included in the bias likelihood model.

Bias likelihood model definition. As noted in section 4.2.1.3, a logistic regression model was used to estimate bias likelihood. The bias likelihood model scores were calculated at the beginning of phases 3 and 4 for the calibration sample and for the main sample (i.e., prior to each intervention) and at the beginning of phases 5 and 6 for the full fielded sample. The bias likelihood model used the current response status for each sample member as its dependent variable each time the bias likelihood model was run.

We label second follow-up survey nonresponse, $y_i^{\text{second follow-up}}$, as 1 for current nonrespondents and 0 for current respondents (as of each time the model is run) and model them with $\Pr(y_i^{\text{second follow-up}}=1)=\text{logit}^{-1}(X_i\beta)$ to reflect the likelihood of contributing to nonresponse bias if remaining a nonrespondent. Input variables are modeled as independent and include school locale (urbanicity), the student's final grade in algebra 1 (algebra), and the remaining covariates specified in table F-3. This model, therefore takes the expanded form

$$\Pr(y_i^{\text{second follow-up}}=1)=\text{logit}^{-1}(\beta_0+\beta_1\bullet\text{urbanicity}_i+\beta_2\bullet\text{algebra}_i+\ldots)$$

From this model, we derive predicted bias likelihood scores, $\hat{p}_i^{\text{bias}}$, for each case, defined as the predicted current nonresponse probability, or

$$\hat{p}_i^{\text{bias}}=\left[\frac{e^{(\beta_0+\beta_1\bullet\text{urbanicity}_i+\beta_2\bullet\text{algebra}_i+\ldots)}}{1+e^{(\beta_0+\beta_1\bullet\text{urbanicity}_i+\beta_2\bullet\text{algebra}_i+\ldots)}}\right]=\Pr(y_i^{\text{second follow-up}}=1)$$

C.3 Calibration Sample and Incentive Experiments

A calibration subsample was selected from each of the three subgroups and was fielded ahead of the main data collection to

experimentally determine optimal incentive amounts for each subgroup. The calibration sample was fielded approximately 8 weeks prior to the main sample to allow time to analyze the experiment results and determine the incentive amounts to be implemented for each subgroup in the main sample. Table C-4 shows the sample size of each subgroup and the number of cases selected for the calibration sample.

Table C-4. Calibration sample sizes, by subgroup

Subgroup	Second follow-up	Calibration sample	Main sample
Total	23,316	3,300	20,016
Subgroup A (high school late/alternative/noncompleters)	2,545	663	1,882
Subgroup B (ultra-cooperative respondents)	4,144	663	3,481
Subgroup C (all other high school completers and unknown cases)	16,627	1,974	14,653

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Second Follow-up.

The calibration sample was fielded in advance of the main sample for the first four of the seven data collection phases used in the second follow-up, after which the calibration and main samples' schedules were synchronized. Table C-5 presents the schedule of data collection phases for both the calibration and main samples. Table C-6 summarizes the baseline and boost incentives tested for each subgroup.

Table C-5. Data collection schedule: 2016

Phase	Calibration sample	Main sample
Phase 1 (baseline incentive)	March 14, 2016	May 9, 2016
Phase 2 (outbound CATI)	March 21, 2016 (subgroup A) and April 4, 2016 (subgroups B and C)	May 16, 2016 (subgroup A) and May 31, 2016 (subgroups B and C)
Phase 3 (incentive boost 1)	May 4, 2016	June 20, 2016
Phase 4 (incentive boost 2)	June 15, 2016	August 1, 2016
Phase 5 (field interviewing)[1]	September 12, 2016	September 12, 2016
Phase 6 (prioritized data collection effort)[1]	November 17, 2016	November 17, 2016
Phase 7 (abbreviated interview)[1]	December 12, 2016	December 12, 2016
End of data collection[1]	January 31, 2017	January 31, 2017

[1] Beginning with phase 5, calibration sample and main sample cases were combined for data collection treatments.
NOTE: Subgroup A = high school late/alternative/noncompleters; subgroup B = ultra-cooperative respondents; subgroup C = all other high school completers and unknown cases; CATI = computer-administered telephone interviewing.
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Second Follow-up.

346 **Table C-6. Baseline and incentive boost experiments for calibration sample: 2016**

Subgroup	Incentive	Amount	Total cumulative incentives offered
Subgroup A (high school late/alternative/noncompleters)	Baseline incentive (all calibration cases)	$0	$0 to $50
		$30	
		$40	
		$50	
	Incentive boost 1 (all remaining calibration nonrespondents)	$15	$15 to $75
		$25	
	Incentive boost 2 (all remaining calibration nonrespondents)	$10	$25 to $95
		$20	
Subgroup B (ultra-cooperative respondents)	Baseline incentive (all calibration cases)	$0	$0 to $50
		$30	
		$40	
		$50	
	Incentive boost 1 (targeted cases only)[1]	$10	$10 to $20 targeted; $0 to $50 otherwise
		$20	
	Incentive boost 2 (targeted cases only)[1]	$10	$10 to $40 targeted; $0 to $50 otherwise
		$20	
Subgroup C (all other high school completers and unknown cases)	Baseline incentive (all calibration cases)	$15	$15 to $40
		$20	
		$25	
		$30	
		$35	
		$40	
	Incentive boost 1 (targeted cases only)	$10	$25 to $60 targeted; $15 to $40 otherwise
		$20	
	Incentive boost 2 (targeted cases only)	$10	$25 to $80 targeted; $15 to $60 otherwise
		$20	

347 [1] Subgroup B (ultra-cooperative respondents) cases offered a nonzero baseline incentive (i.e., $30, $40, or $50) were not eligible to
348 be targeted to receive subsequent treatments (i.e., incentive boost 1 or boost 2).
349 SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009
350 (HSLS:09) Second Follow-up.

C.3.1 Phase 1 and Phase 2 (Baseline Incentive)[5]

During this beginning phase of data collection, the survey was open exclusively for self-administered interviews via the web (except for instances when sample members called into the study help desk) and no outbound telephone prompting occurred. Calibration sample members were randomized to different incentive levels within subgroups to identify the optimal baseline amounts to be offered to main sample cases.

After phase 1, telephone interviewers began making outbound calls to prompt sample members to complete the interview over the telephone or by web-based self-administration, as part of phase 2. Outbound computer-assisted telephone interviewing (CATI) began earlier for cases in subgroup A (HSNC) to allow additional time for telephone interviewers to work these high-priority cases. No additional incentives were offered during phase 2.

To assess the efficacy of the baseline incentive amounts offered, chi-square tests were used to perform pairwise comparisons between response rates by incentive levels within each of the three subgroups. Results of these comparisons are shown below for each subgroup.

Subgroup A (HSNC). Table F-7 displays subgroup A response rates by baseline incentive level. About 6 percent of cases in subgroup A who did not receive an incentive offer responded by the end of phase 2. Among this set of cases, unincentivized (i.e., $0 incentive) cases were significantly less likely to respond compared to the next lowest incentive level of $30 ($\chi 2$ (1, N = 324) = 18.72, p < .05). Response rates were highest among cases assigned a baseline

[5] The calibration HSNC (subgroup A) subsample was intended to receive a baseline incentive offer ($30, $40, or $50) whereas calibration UC (subgroup B) cases were intended not to be offered a baseline incentive. In the original selection of calibration cases, the subgroup A cases and subgroup B cases were misclassified such that 154 subgroup A cases were not offered a baseline incentive while 509 subgroup B cases were offered a baseline incentive ($30, $40, or $50). Upon discovery of this error, 509 additional HSNC and 154 additional UC cases were redrawn for the calibration sample and given an incentive offer (or no incentive offer) as originally intended. The misclassified cases continued to be worked throughout the remainder of data collection, although the incentivized subgroup B cases were not eligible to receive additional incentive boosts.

incentive of $40 (29 percent). The $40 response rate is about 6 percentage points higher than the $30 rate (23 percent), although not significantly higher at the 0.05 level, ($\chi 2$ (1, N = 340) = 1.84, p = .17). No significant difference was detected between response rates at the $40 incentive level and the $50 level. Given the magnitude of the observed difference between $30 and $40, a baseline incentive of $40 was offered to all cases in the subgroup A main sample.

Table C-7. Subgroup A response rates by baseline incentive amount as of April 27, 2016

Baseline incentive offer	Sample members (*n*)	Respondents (*n*)	Response rate (percent)
Total	**663**	**147**	**22.2**
$0	154	9	5.8
$30	170	39	22.9
$40	170	50	29.4
$50	169	49	29.0

NOTE: Excludes partially completed cases.
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Second Follow-up.

Subgroup B (UC). Table F-8 displays subgroup B response rates, after approximately 5 weeks of data collection, by baseline incentive level. For context, table C-9 presents subgroup B response rates together with response rates for other selected NCES studies. The selected studies include the 2012/14 Beginning Postsecondary Students Longitudinal Study (BPS:12/14), as the BPS:12/14 and HSLS:09 second follow-up sample members are similar in age, and the 2008/12 Baccalaureate and Beyond Longitudinal Study (B&B:08/12), as these sample members are another highly cooperative population. The results shown in table F-9 indicate that the HSLS:09 subgroup of ultra-cooperative calibration sample members responded, with no incentive offer, at a rate similar to that seen among BPS:12/14 calibration sample members with high predicted response likelihood and with a $40 incentive (after 5 weeks of data collection). The unincentivized ultra-cooperative calibration sample response rate of 64 percent is also similar to that seen among B&B:08/12 sample members who had responded during the early response period (i.e., after 4 weeks of data collection) of B&B:08/12 and its first follow-up round of data collection. Given the strong response rate for subgroup B, no baseline incentive was offered to subgroup B cases in the main sample.

Table C-8. Subgroup B response rates by baseline incentive amount as of April 27, 2016

Baseline incentive offer	Sample members (*n*)	Respondents (*n*)	Response rate (percent)
Total	**663**	**493**	**74.4**
$0	154	98	63.6
$30	170	127	74.7
$40	170	134	78.8
$50	169	134	79.3

NOTE: Excludes partially completed cases.
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Second Follow-up.

Table C-9. Comparison of subgroup B response rates with response rates from selected studies

Study group	Response rate (percent)
HSLS:09 second follow-up calibration sample (subgroup B, phases 1 and 2)[1]	
No baseline incentive offer	63.6
$30 baseline incentive offer	74.7
$40 baseline incentive offer	78.8
$50 baseline incentive offer	79.3
BPS:12/14 calibration sample (response likelihood > .9, after 5 weeks)	
No incentive offer	23.5
$10 incentive offer	29.6
$20 incentive offer	43.9
$30 incentive offer	58.8
$40 incentive offer	61.9
$50 incentive offer	66.3
B&B:08/12 early response phase[2] respondents, by prior round response status	
Base year (NPSAS:08) and first follow-up (B&B:08/09) respondents	48.1
First follow-up (B&B:08/09) early response phase[2] respondents	64.5
Base year (NPSAS:08) and first follow-up (B&B:08/09) early response phase[2] respondents	69.9

[1] Excludes partially completed cases.
[2] The B&B:08/08 and the B&B:08/12 early response phases consisted of the first 4 weeks of data collection.
NOTE: HSLS:09 = High School Longitudinal Study of 2009; BPS:12/14 = 2012/14 Beginning Postsecondary Students Longitudinal Study; B&B:08/12 = 2008/12 Baccalaureate and Beyond Longitudinal Study; BPS:08/09 = 2008/2009 Beginning Postsecondary Students Longitudinal Study; NPSAS:08 = 2007–08 National Postsecondary Student Aid Study.
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Second Follow-up; U.S. Department of Education, National Center for Education Statistics, 2012/14 Beginning Postsecondary Students Longitudinal Study (BPS:12/14); U.S. Department of Education, National Center for Education Statistics, 2008/12 Baccalaureate and Beyond Longitudinal Study (B&B:08/12).

Subgroup C (HS other). Table F-10 provides subgroup C (HS other) response rates by baseline incentive level. Within subgroup C, the highest response rate, 43 percent, was observed among cases assigned a $30 incentive. No significant difference was detected

between the response rate associated with the $30 baseline incentive and that of either the $35 incentive or $40 incentive. Response rates among cases assigned the $30 incentive were significantly higher than those for $15 and $20 ($\chi 2$ (1, N = 658) = 17.28, $p < .05$ and $\chi 2$ (1, N = 658) = 6.59, $p < .05$, respectively).

No significant difference was detected at the .05 level between comparisons of response rates for cases assigned $30 (43 percent) and $25 (37 percent) ($\chi 2$ (1, N = 658) = 2.53, p = .11). Given that subgroup C constitutes the largest subgroup in the main sample, with more than 14,000 sample members, a 6 percent difference in response rate would result in a nontrivial difference in yield; as such, a baseline incentive of $30 was offered to all subgroup C main sample cases.

Table C-10. Subgroup C response rates by baseline incentive amount as of April 27, 2016

Baseline incentive offer	Sample members (*n*)	Respondents (*n*)	Response rate (percent)
Total	**1,974**	**733**	**37.1**
$15	329	91	27.7
$20	329	110	33.4
$25	329	122	37.1
$30	329	142	43.2
$35	329	130	39.5
$40	329	138	41.9

NOTE: Excludes partially completed cases.
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Second Follow-up.

C.3.2 Phase 3 (Incentive Boost 1 Offer)

Phase 3 of the calibration study introduced an incentive boost that was offered to a subset of pending nonrespondents in addition to the baseline amount offered in the prior phases. The bias likelihood model was deployed prior to the start of phase 3 and was used to target subgroup B and subgroup C cases to receive an incentive boost (boost 1) in addition to their baseline incentive, should they complete the survey. Given the relative importance of obtaining responses from subgroup A cases, all remaining nonrespondent cases in subgroup A were targeted for an incentive boost offer.

Subgroup A (HSNC). Table F-11 displays subgroup A response rates during phase 3 by incentive boost level and baseline incentive level. For subgroup A cases that received no baseline incentive, no significant difference was detected between the response rates of sample members who were offered the $15 (10 percent) and $25 (15 percent) boost 1 incentive. No significant differences were detected between the response rates of sample members who were offered the $15 (17 percent) and $25 (12 percent) boost 1 incentive, when the baseline incentive was $30. Additionally, there was no significant difference detected between the response rates of sample members who were offered the $15 (12 percent) and $25 (19 percent) boost 1 incentive, when the baseline incentive was $40. Lastly, no significant differences were detected between the response rates of sample members who were offered the $15 (12 percent) and $25 (17 percent) boost 1 incentive, when the baseline incentive was $50. Given that no significant differences were found between the $15 and $25 boost incentives, based on the results available on June 7, 2016, a boost 1 incentive of $15 was offered to all phase 3 cases in the subgroup A main sample.

Table C-11. Subgroup A response rates in phase 3, by boost 1 incentive amount as of June 7, 2016

Boost 1 incentive offer	Sample members (*n*)	Respondents (*n*)	Response rate (percent)
Total	509	71	13.9
No baseline incentive, $15 boost	73	7	9.6
No baseline incentive, $25 boost	72	11	15.3
Baseline incentive, $15 boost	185	25	13.5
$30 Baseline incentive	66	11	16.7
$40 Baseline incentive	**59**	**7**	**11.9**
$50 Baseline incentive	60	7	11.7
Baseline incentive, $25 boost	179	28	15.6
$30 Baseline incentive	61	7	11.5
$40 Baseline incentive	**58**	**11**	**19.0**
$50 Baseline incentive	60	10	16.7

NOTE: Excludes partially completed cases. Bolded text indicates the baseline incentive offered to the main sample.
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Second Follow-up.

Subgroup B (UC). Table F-12 presents response rates during phase 3 by incentive boost level for subgroup B cases targeted by the bias likelihood model for intervention. Note that most of the ultra-cooperative sample members had previously responded in phases 1 and 2, leaving very few nonrespondents eligible to be targeted for an incentive intervention in phase 3 (18 targeted cases). Additionally, subgroup B sample members assigned a nonzero baseline incentive were not targeted for boost 1 incentives. Given the small number of cases within subgroup B, statistical analysis of the boost 1 incentive was not conducted, and the minimum incentive ($10) was offered to all phase 3 targeted subgroup B main sample cases.

Table C-12. Subgroup B response rates in phase 3, by boost 1 incentive amount as of June 7, 2016

Boost 1 incentive offer	Sample members (*n*)	Respondents (*n*)	Response rate (percent)
Total	**18**	**5**	**27.8**
No baseline incentive, $10 boost	9	3	33.3
No baseline incentive, $20 boost	9	2	22.2

NOTE: Excludes partially completed cases and subgroup B cases offered a nonzero baseline incentive (i.e., $30, $40, or $50).
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Second Follow-up.

Subgroup C (HS other). Table C-13 displays subgroup C response rates during phase 3 by incentive level, among the 661 cases selected for an incentive boost offer based on the bias likelihood model. No significant difference was detected between the phase 3 response rates of sample members offered $10 (13.9 percent) and $20 (15.5 percent) boost 1 incentives, regardless of the baseline incentive offered. As such, a boost 1 incentive of $10 was offered to all phase 3 targeted cases in the subgroup C main sample.

Table C-13. Subgroup C response rates in phase 3, by boost 1 incentive amount as of June 7, 2016

Boost 1 incentive offer	Sample members (*n*)	Respondents (*n*)	Response rate (percent)
Total	**661**	**97**	**14.7**
Baseline incentive, $10 boost	332	46	13.9
$15 Baseline incentive	64	8	12.5
$20 Baseline incentive	58	6	10.3
$25 Baseline incentive	54	7	13.0
$30 Baseline incentive	**45**	**6**	**13.3**
$35 Baseline incentive	55	7	12.7
$40 Baseline incentive	56	12	21.4
Baseline incentive, $20 boost	329	51	15.5
$15 Baseline incentive	61	9	14.8
$20 Baseline incentive	61	5	8.2
$25 Baseline incentive	52	12	23.1
$30 Baseline incentive	**46**	**8**	**17.4**
$35 Baseline incentive	53	9	17.0
$40 Baseline incentive	56	8	14.3

NOTE: Excludes partially completed cases. Bolded text indicates the baseline incentive offered to the main sample.
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Second Follow-up.

C.3.3 Phase 4 (Incentive Boost 2 Offer and Adaptive Incentive Boost 2b Offer)

Phase 4 of the calibration study introduced a second incentive boost that was offered to a subset of pending nonrespondents in addition to the baseline amount and first boost, as applicable. The bias likelihood model was deployed again prior to the start of phase 4 and was again used to identify cases in subgroup B and subgroup C for targeted interventions (i.e., to receive an incentive boost offer). Note that cases were selected for the boost 2 offer independently from the selection of cases for boost 1. A case targeted for a boost 1 incentive offer might or might not be selected to receive a boost 2 incentive offer depending on how its bias likelihood score shifted between the phases. As was done in phase 3, all remaining nonrespondent cases in subgroup A were targeted for an incentive boost 2 offer. An initial analysis of the boost 2 incentive was conducted after 4 weeks (July 15, 2016) to determine the optimal incentive amount for the main sample. However, a second analysis after approximately 11 weeks (September 7, 2016) revealed that the results had shifted for subgroups A and C, as detailed below.

Subgroup A (HSNC). Results for the boost 2 incentive offer for subgroup A, assessed after 4 weeks, are presented in table F-14. No significant differences were detected between response rates among cases assigned the $10 and $20 boost incentives. Due to the small number of respondents in phase 4, results are not disaggregated by baseline or boost 1 incentive levels. Therefore, a boost 2 of $10 was initially selected for subgroup A main sample cases.

Subgroup B (UC). Results for the boost 2 incentive for subgroup B are presented in table F-15. As with boost 1, subgroup B sample members assigned a nonzero baseline incentive were not targeted for boost 2 incentives. No statistical comparisons were performed due to the small number of cases in this condition. A boost 2 of $10 was selected for subgroup B main sample cases.

Subgroup C (HS other). Results for the boost 2 incentive for subgroup C are presented in table F-16. Like subgroup A and subgroup B, due to the small number of respondents in phase 4, results are not disaggregated by previous baseline or boost 1 incentive levels. No significant differences in response rates were found between cases assigned the $10 and $20 boost levels. As such, a boost 2 of $10 was initially selected for subgroup C main sample cases.

Incentive boost 2b. While response rates for cases assigned to $10 and $20 boost 2 incentive levels were statistically equivalent (i.e., no significant differences were detected) at 4 weeks for each of the subgroups, when reassessed after about 11 weeks (September 7, 2016) the differences between cases assigned $10 and $20 had become large and statistically significant for subgroup A (χ^2 (1, N = 310) = 6.38, p < .05) and subgroup C (χ^2 (1, N = 576) = 4.02, p < .05). (Subgroup B had very small numbers and no detectable difference.) The additional time for the calibration sample cases in phase 4 revealed an effect that was not evident at the end of the first 4 weeks of phase 4. In the intervening weeks, staff increased locating, prompting, and case review efforts for all pending cases (regardless of incentive amount assignment). Results after 4 weeks in phase 4 and after 11 weeks in phase 4 are presented below in tables F-14, F-15, and F-16.

Table C-14. Subgroup A phase 4 calibration results after 4 weeks and after 11 weeks, by boost 2 incentive amount: 2016

		Boost 2 results after 4 weeks		Boost 2 results after 11 weeks	
Boost 2 incentive offer	Sample members (*n*)	Respondents (*n*)	Response rate (percent)	Respondents (*n*)	Response rate (percent)
Total	**310**	**17**	**5.5**	**39**	**12.6**
$10	154	8	5.2	12	7.8
$20	156	9	5.8	27	17.3

NOTE: Excludes partially completed cases.
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Second Follow-up.

Table C-15. Subgroup B phase 4 calibration results after 4 weeks and after 11 weeks, by boost 2 incentive amount: 2016

		Boost 2 results after 4 weeks		Boost 2 results after 11 weeks	
Boost 2 incentive offer	Sample members (*n*)	Respondents (*n*)	Response rate (percent)	Respondents (*n*)	Response rate (percent)
Total	**14**	**2**	**14.3**	**4**	**28.6**
$10	7	1	14.3	2	28.6
$20	7	1	14.3	2	28.6

NOTE: Excludes partially completed cases and subgroup B cases offered a nonzero baseline incentive (i.e., $30, $40, or $50).
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Second Follow-up.

Table C-16. Subgroup C phase 4 calibration results after 4 weeks and after 11 weeks, by boost 2 incentive amount: 2016

		Boost 2 results after 4 weeks		Boost 2 results after 11 weeks	
Boost 2 incentive offer	Sample members (*n*)	Respondents (*n*)	Response rate (percent)	Respondents (*n*)	Response rate (percent)
Total	**576**	**36**	**6.3**	**81**	**14.1**
$10	287	17	5.9	32	11.1
$20	289	19	6.6	49	17.0

NOTE: Excludes partially completed cases.
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Second Follow-up.

Based on results after 11 weeks in phase 4, an adaptive component was added to the responsive design protocol in which an additional

boost (incentive boost 2b) of $10 was offered to subgroup A main sample nonrespondents and subgroup C main sample boost 2-targeted cases; no additional boost was offered to subgroup B cases.

C.4 Assessment of Responsive Design Models

This section provides an assessment of the effectiveness and results of the response likelihood model and bias likelihood model.

C.4.1 Assessment of Response Likelihood Model on Second Follow-up Response Rates

As noted previously, the response likelihood model was fit once, prior to the start of the second follow-up data collection, and was designed to predict the likelihood of a case becoming a respondent. To assess the performance of the response likelihood model on realized response rates, response likelihood scores (predicted probabilities from the response likelihood logistic regression model) were ordered into deciles and response rates were examined within those deciles. Deciles were created using the SAS RANK procedure which defaults to placing cases with identical values into the higher ranked category, thereby preventing any two deciles including the same predicted probabilities. Table F-17 shows response rates by response likelihood decile.

Table C-17. Response rates by response likelihood score deciles: 2016

Response likelihood decile	Sample members[1] (*n*)	Respondents	Response rate
Total	**23,316**	**17,335**	**74.3**
1	2,332	1,027	44.0
2	2,333	1,239	53.1
3	2,329	1,614	69.3
4	2,341	1,785	76.2
5	2,319	1,806	77.9
6	2,395	1,926	80.4
7	2,194	1,778	81.0
8	2,471	2,065	83.6
9	2,237	1,970	88.1
10	2,365	2,125	89.9

[1] Note the total sample (23,316) represents to total fielded sample and excludes sample members that withdrew from the study between the end of the 2013 Update collection and the beginning of the second follow-up data collection or were found to be deceased.
SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Second Follow-up.

Second follow-up response rates increased as the predicted response probability decile increased, indicating that a higher predicted response likelihood was associated with a higher likelihood of becoming a study respondent. The general pattern across all deciles indicates that the response likelihood model was effective in ordinally predicting a case's response outcome.

C.4.2 Assessment of Bias Likelihood Model on Sample Representativeness

As described in section 4.2.1.3, the bias likelihood model was used to identify cases that were most unlike the set of sample members that had responded at each time-point the model was fit. The model used key survey and frame variables as model covariates with current nonresponse (as of each model run) as the dependent variable to identify nonrespondents most likely to contribute to bias in key survey variables unless converted to respondents. The bias likelihood model was fit at the beginning of phases 3 and 4 for the calibration and main samples (i.e., prior to both boost interventions) and at the beginning of phases 5 and 6[6] for the combined sample.

To assess the effectiveness of the bias likelihood model on sample representativeness, *weighted* estimates of key model variables were examined at baseline (i.e., for all sample members) and then throughout the phases of data collection. Weighted estimates were examined to provide information on the values of these important variables in the population of interest, rather than in the sample. Table F-18 shows the weighted estimates of the key analytic variables used in the bias likelihood model at baseline and at the time of selection of targeted cases for each phase.

[6] Beginning with phase 5, calibration sample and main sample cases were combined for data collection treatments. Note that phases 5 and 6 were not part of the calibration experiment, and are therefore not covered in this appendix. For details on these phases, see section 4.2.1.4.

633 **Table C-18. Weighted estimates of bias likelihood model variables and other key variables, at baseline, phase target selection,**
634 **and data collection end**

	Baseline		Phase 3			Phase 4			Phase 5			Phase 6			Data Collection End	
Domain category	*n*	%	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	*n*	%
School Type																
Public	3,007,154	92.95	1,023,314	91.53	94.70	1,348,003	91.83	95.39	1,604,809	91.97	97.43	1,845,884	92.31	94.81	2,177,263	92.43
Catholic	120,717	3.73	53,727	4.81	2.39	66,810	4.55	1.57	76,913	4.41	0.99	82,854	4.14	2.64	94,556	4.01
Other private	107,318	3.32	40,937	3.66	2.92	53,177	3.62	3.04	63,222	3.62	1.59	70,936	3.55	2.55	83,811	3.56
Sex																
Male	1,634,337	50.52	472,687	42.28	70.16	667,454	45.47	60.22	801,376	45.93	66.60	942,856	47.15	57.24	1,124,667	47.74
Female	1,600,852	49.48	645,291	57.72	29.84	800,537	54.53	39.78	943,568	54.07	33.40	1,056,819	52.85	42.76	1,230,963	52.26
Race/ethnicity[1]																
American Indian / Alaska Native / Native Hawaiian / Pacific Islander	39,093	1.21	10,819	0.97	0.87	13,181	0.90	1.46	16,662	0.95	1.81	19,261	0.96	1.90	24,366	1.03
Hispanic	721,720	22.31	220,775	19.75	30.85	308,906	21.04	24.41	374,515	21.46	19.92	430,535	21.53	24.76	507,575	21.55
Asian	116,583	3.60	46,834	4.19	3.81	61,583	4.20	2.33	72,708	4.17	0.58	79,360	3.97	2.58	90,350	3.84
Black	437,312	13.52	130,779	11.70	14.11	173,042	11.79	16.14	204,000	11.69	32.59	256,686	12.84	15.02	306,216	13.00
More than one race	240,128	7.42	71,840	6.43	8.85	99,331	6.77	10.43	128,424	7.36	7.31	148,540	7.43	7.51	175,419	7.45
White	1,680,353	51.94	636,931	56.97	41.50	811,947	55.31	45.23	948,635	54.36	37.79	1,065,294	53.27	48.23	1,251,703	53.14
School locale (urbanicity)																
City	947,003	29.27	331,594	29.66	34.46	441,948	30.11	29.72	525,903	30.14	30.56	604,255	30.22	27.49	702,039	29.80
Suburb	899,197	27.79	315,818	28.25	26.23	413,595	28.17	25.60	486,237	27.87	29.61	561,049	28.06	27.48	661,567	28.08
Town	416,617	12.88	136,153	12.18	10.56	177,404	12.08	14.54	214,697	12.30	10.11	240,950	12.05	14.17	291,954	12.39
Rural	972,372	30.06	334,413	29.91	28.75	435,044	29.64	30.13	518,107	29.69	29.71	593,420	29.68	30.86	700,070	29.72

See notes at end of table.

Table C-18. Weighted estimates of bias likelihood model variables and other key variables, at baseline, phase target selection, and data collection end—Continued

Domain category	Baseline		Phase 3			Phase 4			Phase 5			Phase 6			Data Collection End	
	n	%	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	*n*	%
Teenager's final grade in algebra I																
A	1,073,268	33.17	456,321	40.82	21.79	571,617	38.94	19.36	660,319	37.84	17.73	722,910	36.15	27.27	831,177	35.28
B	1,157,212	35.77	368,499	32.96	37.57	493,575	33.62	43.01	595,674	34.14	37.50	699,909	35.00	36.86	824,123	34.99
C	659,894	20.40	195,699	17.50	24.07	265,450	18.08	25.65	327,458	18.77	31.09	385,060	19.26	23.04	465,978	19.78
D or lower	262,124	8.10	72,319	6.47	14.63	105,597	7.19	9.39	124,537	7.14	8.73	146,179	7.31	9.60	180,025	7.64
Ungraded / have not completed class	82,691	2.56	25,139	2.25	1.93	31,752	2.16	2.60	36,957	2.12	4.95	45,617	2.28	3.23	54,325	2.31
How far in school 9th-grader thinks he/she will go																
High school graduate or less	472,264	14.60	112,213	10.04	19.09	160,545	10.94	21.41	198,202	11.36	33.04	255,813	12.79	18.87	315,083	13.38
Some college	241,892	7.48	69,443	6.21	9.67	97,869	6.67	10.59	122,451	7.02	7.31	141,355	7.07	8.29	167,209	7.10
College graduate	554,233	17.13	213,117	19.06	13.96	275,485	18.77	16.25	325,406	18.65	10.02	361,714	18.09	14.37	415,768	17.65
Master's degree	646,291	19.98	250,802	22.43	18.30	324,069	22.08	16.65	374,937	21.49	10.83	415,883	20.80	17.67	486,445	20.65
Doctor's degree	613,655	18.97	235,581	21.07	20.20	308,623	21.02	14.84	370,031	21.21	9.60	410,395	20.52	16.13	471,498	20.02
Don't know	706,854	21.85	236,822	21.18	18.78	301,399	20.53	20.27	353,918	20.28	29.20	414,515	20.73	24.68	499,626	21.21

See notes at end of table.

57 **Table C-18. Weighted estimates of bias likelihood model variables and other key variables, at baseline, phase target selection,**
58 **and data collection end—Continued**

	Baseline		Phase 3			Phase 4			Phase 5			Phase 6			Data Collection End	
Domain category	*n*	%	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	*n*	%
How far in school parent thinks 9th-grader will go																
High school graduate or less	319,438	9.87	76,373	6.83	11.08	103,703	7.06	12.89	124,296	7.12	21.56	158,267	7.91	14.23	201,729	8.56
Some college	332,596	10.28	92,116	8.24	12.19	124,434	8.48	14.87	151,921	8.71	21.15	190,587	9.53	12.09	227,963	9.68
College graduate	935,916	28.93	344,961	30.86	26.51	448,437	30.55	27.68	530,266	30.39	18.01	594,927	29.75	26.31	688,892	29.24
Master's degree	610,813	18.88	236,404	21.15	19.78	314,166	21.40	12.47	368,719	21.13	7.45	401,538	20.08	16.70	468,468	19.89
Doctor's degree	661,154	20.44	251,271	22.48	17.04	320,683	21.85	17.90	381,352	21.85	19.16	434,109	21.71	18.42	500,540	21.25
Don't know	375,273	11.60	116,853	10.45	13.40	156,568	10.67	14.19	188,391	10.80	12.67	220,247	11.01	12.25	268,036	11.38
How far in school sample member thinks he/she will go																
High school graduate or less	560,041	17.31	145,399	13.01	21.41	199,524	13.59	22.72	239,672	13.74	33.03	294,729	14.74	22.98	362,565	15.39
Some college	375,268	11.60	112,648	10.08	13.06	151,040	10.29	14.32	183,869	10.54	14.95	211,880	10.60	13.93	262,817	11.16
College graduate	899,602	27.81	325,828	29.14	32.13	436,090	29.71	25.13	514,611	29.49	21.22	582,519	29.13	24.45	673,694	28.60
Master's degree	653,917	20.21	264,764	23.68	14.24	336,427	22.92	15.05	399,320	22.88	12.82	440,446	22.03	16.50	506,506	21.50
Doctor's degree	391,499	12.10	161,066	14.41	8.57	200,647	13.67	9.97	234,405	13.43	3.61	267,852	13.39	9.09	306,256	13.00
Don't know	354,862	10.97	108,272	9.68	10.58	144,263	9.83	12.81	173,067	9.92	14.37	202,248	10.11	13.05	243,790	10.35

See notes at end of table.

59
60 **Table C-18. Weighted estimates of bias likelihood model variables and other key variables, at baseline, phase target selection, and data collection end—Continued**

	Baseline		Phase 3			Phase 4			Phase 5			Phase 6			Data Collection End	
Domain category	*n*	%	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	*n*	%
How far in school parent thinks sample member will go																
High school graduate or less	486,717	15.04	142,986	12.79	18.96	198,231	13.50	17.93	235,180	13.48	21.56	282,231	14.11	16.98	339,606	14.42
Some college	334,677	10.34	103,051	9.22	9.65	134,880	9.19	11.12	159,971	9.17	16.54	193,150	9.66	12.17	232,264	9.86
College graduate	968,389	29.93	343,589	30.73	31.61	454,749	30.98	25.53	540,208	30.96	23.54	605,843	30.30	29.01	712,360	30.24
Master's degree	579,701	17.92	223,998	20.04	16.51	292,477	19.92	15.57	347,058	19.89	11.30	388,886	19.45	15.36	451,608	19.17
Doctor's degree	463,243	14.32	181,734	16.26	11.25	228,935	15.60	14.46	270,807	15.52	9.73	304,400	15.22	11.88	348,169	14.78
Don't know	402,461	12.44	122,620	10.97	12.02	158,719	10.81	15.38	191,722	10.99	17.32	225,164	11.26	14.59	271,621	11.53
Grade level in spring 2012 or last date of attendance																
9th or 10th grade	83,441	2.58	22,139	1.98	3.13	29,638	2.02	2.67	33,365	1.91	4.66	42,237	2.11	3.66	52,426	2.23
11th grade	2,958,759	91.46	1,046,440	93.60	92.30	1,377,197	93.82	89.39	1,631,816	93.52	80.64	1,854,641	92.75	87.95	2,174,033	92.29
12th grade	112,609	3.48	30,207	2.70	2.63	37,001	2.52	4.96	49,549	2.84	7.58	61,870	3.09	4.58	75,944	3.22
Ungraded program	14,957	0.46	5,295	0.47	0.22	5,855	0.40	0.37	6,435	0.37	1.52	8,264	0.41	0.59	10,712	0.45
Not attending high school during 2011–12 school year	65,423	2.02	13,897	1.24	1.72	18,300	1.25	2.61	23,779	1.36	5.61	32,662	1.63	3.21	42,515	1.80

See notes at end of table.

61 **Table C-18. Weighted estimates of bias likelihood model variables and other key variables, at baseline, phase target selection,**
62 **and data collection end—Continued**

	Baseline		Phase 3			Phase 4			Phase 5			Phase 6			Data Collection End	
Domain category	*n*	%	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	*n*	%
Student dual first language indicator																
First language is English only	2,668,349	82.48	933,194	83.47	77.53	1,215,570	82.81	82.18	1,441,246	82.60	86.84	1,654,199	82.72	81.78	1,950,799	82.81
First language is non-English only	374,115	11.56	114,836	10.27	16.83	163,250	11.12	12.05	195,461	11.20	10.43	226,477	11.33	12.05	265,110	11.25
First language is English and non-English	192,725	5.96	69,949	6.26	5.64	89,169	6.07	5.78	108,237	6.20	2.72	118,998	5.95	6.17	139,721	5.93
9th-grader is taking math course in fall 2009 term																
No	324,809	10.04	88,641	7.93	14.22	125,897	8.58	13.31	154,894	8.88	12.86	182,533	9.13	11.99	222,626	9.45
Yes	2,910,380	89.96	1,029,336	92.07	85.78	1,342,093	91.42	86.69	1,590,051	91.12	87.14	1,817,141	90.87	88.01	2,133,004	90.55
9th-grader is taking science course in fall 2009 term																
No	580,257	17.94	168,640	15.08	22.20	231,033	15.74	22.80	279,616	16.02	26.02	329,992	16.50	20.67	401,122	17.03
Yes	2,654,932	82.06	949,338	84.92	77.80	1,236,957	84.26	77.20	1,465,329	83.98	73.98	1,669,682	83.50	79.33	1,954,508	82.97
Attended career day or job fair																
No	1,672,362	51.69	585,001	52.33	54.74	768,946	52.38	51.40	912,402	52.29	52.61	1,041,006	52.06	50.62	1,221,717	51.86
Yes	1,562,827	48.31	532,977	47.67	45.26	699,045	47.62	48.60	832,543	47.71	47.39	958,668	47.94	49.38	1,133,913	48.14
Attended program at or took tour of college campus																
No	1,586,649	49.04	513,462	45.93	50.52	678,338	46.21	50.31	810,657	46.46	58.77	940,505	47.03	54.02	1,120,284	47.56
Yes	1,648,540	50.96	604,516	54.07	49.48	789,653	53.79	49.69	934,287	53.54	41.23	1,059,170	52.97	45.98	1,235,346	52.44

63 **Table C-18. Weighted estimates of bias likelihood model variables and other key variables, at baseline, phase target selection,**
64 **and data collection end—Continued**

	Baseline		Phase 3			Phase 4			Phase 5			Phase 6			Data Collection End	
Domain category	*n*	%	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	*n*	%
See notes at end of table.																
Repeated grade																
No	3,031,677	93.71	1,053,164	94.20	93.27	1,384,038	94.28	93.82	1,646,121	94.34	90.85	1,884,315	94.23	92.29	2,213,191	93.95
Yes	203,512	6.29	64,814	5.80	6.73	83,953	5.72	6.18	98,824	5.66	9.15	115,359	5.77	7.71	142,439	6.05
Sat in on or took college class																
No	2,410,326	74.50	796,871	71.28	78.62	1,063,383	72.44	77.44	1,266,706	72.59	82.02	1,458,016	72.91	77.44	1,730,899	73.48
Yes	824,862	25.50	321,107	28.72	21.38	404,608	27.56	22.56	478,238	27.41	17.98	541,658	27.09	22.56	624,731	26.52
Participated in internship or apprenticeship related to career goals																
No	2,704,701	83.60	955,413	85.46	80.62	1,244,812	84.80	80.63	1,478,556	84.73	74.78	1,681,671	84.10	82.31	1,977,167	83.93
Yes	530,488	16.40	162,565	14.54	19.38	223,178	15.20	19.37	266,389	15.27	25.22	318,004	15.90	17.69	378,464	16.07
Performed paid/volunteer work in job related to career goals																
No	2,136,745	66.05	753,875	67.43	67.88	987,985	67.30	65.23	1,171,033	67.11	66.74	1,339,170	66.97	64.78	1,564,290	66.41
Yes	1,098,443	33.95	364,103	32.57	32.12	480,005	32.70	34.77	573,912	32.89	33.26	660,505	33.03	35.22	791,340	33.59
Searched Internet or read college guides for college options																
No	646,273	19.98	181,737	16.26	23.80	247,005	16.83	23.45	292,824	16.78	27.39	350,260	17.52	25.46	431,026	18.30
Yes	2,588,916	80.02	936,241	83.74	76.20	1,220,986	83.17	76.55	1,452,120	83.22	72.61	1,649,415	82.48	74.54	1,924,604	81.70

65 **Table C-18. Weighted estimates of bias likelihood model variables and other key variables, at baseline, phase target selection,**
66 **and data collection end—Continued**

	Baseline		Phase 3			Phase 4			Phase 5			Phase 6			Data Collection End	
Domain category	*n*	%	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	*n*	%
Talked w/ high school counselor about options for after high school																
No	1,199,704	37.08	410,941	36.76	36.46	535,295	36.46	39.83	645,845	37.01	39.16	739,719	36.99	37.47	875,322	37.16
Yes	2,035,485	62.92	707,037	63.24	63.54	932,695	63.54	60.17	1,099,099	62.99	60.84	1,259,955	63.01	62.53	1,480,309	62.84

See notes at end of table.

Talked about options w/ counselor hired to prepare for college admission																
No	2,832,193	87.54	989,473	88.51	85.49	1,293,087	88.09	88.58	1,541,517	88.34	84.80	1,768,357	88.43	86.22	2,070,848	87.91
Yes	402,996	12.46	128,505	11.49	14.51	174,903	11.91	11.42	203,428	11.66	15.20	231,317	11.57	13.78	284,782	12.09
Took course to prepare for college admission exam																
No	1,936,450	59.86	637,944	57.06	62.17	842,142	57.37	63.56	1,011,442	57.96	62.52	1,164,338	58.23	64.27	1,379,843	58.58
Yes	1,298,739	40.14	480,034	42.94	37.83	625,849	42.63	36.44	733,503	42.04	37.48	835,337	41.77	35.73	975,787	41.42
Teenager taking math class(es) in spring 2012																
No	465,128	14.38	134,842	12.06	16.32	178,707	12.17	17.01	213,083	12.21	28.37	259,973	13.00	17.07	314,935	13.37
Yes	2,770,061	85.62	983,136	87.94	83.68	1,289,283	87.83	82.99	1,531,861	87.79	71.63	1,739,702	87.00	82.93	2,040,695	86.63

Table C-18. Weighted estimates of bias likelihood model variables and other key variables, at baseline, phase target selection, and data collection end—Continued

Domain category	Baseline		Phase 3			Phase 4			Phase 5			Phase 6			Data Collection End	
	n	%	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	*n*	%
Sample member has high school credential																
No	402,808	12.45	106,479	9.52	10.95	138,178	9.41	14.28	162,597	9.32	29.88	205,999	10.30	17.18	260,443	11.06
Yes	2,832,380	87.55	1,011,499	90.48	89.05	1,329,812	90.59	85.72	1,582,348	90.68	70.12	1,793,675	89.70	82.82	2,095,187	88.94
Taking postsecondary classes as of Nov. 1, 2013																
Yes	2,175,181	67.24	849,917	76.02	59.94	1,099,243	74.88	54.28	1,290,075	73.93	38.01	1,444,087	72.22	55.98	1,658,467	70.40
No	685,990	21.20	171,657	15.35	28.52	241,240	16.43	30.52	300,321	17.21	36.37	361,488	18.08	27.76	452,932	19.23
Don't know	374,018	11.56	96,404	8.62	11.54	127,507	8.69	15.20	154,548	8.86	25.62	194,100	9.71	16.25	244,230	10.37

See notes at end of table.

Level of program enrolled in as of Nov. 1, 2013																
Bachelor's degree	1,200,395	37.10	517,218	46.26	26.29	653,967	44.55	21.71	759,748	43.54	7.57	831,692	41.59	27.62	938,948	39.86
Associate's degree	464,242	14.35	168,589	15.08	13.95	226,023	15.40	11.85	264,683	15.17	11.62	301,250	15.06	12.79	348,568	14.80
Certificate or diploma program from school that provides occupational training	102,564	3.17	33,801	3.02	4.83	46,267	3.15	4.34	53,729	3.08	2.79	62,068	3.10	2.82	71,586	3.04
Other	1,467,988	45.38	398,369	35.63	54.93	541,732	36.90	62.11	666,785	38.21	78.03	804,664	40.24	56.77	996,527	42.30

69
70

Table C-18. Weighted estimates of bias likelihood model variables and other key variables, at baseline, phase target selection, and data collection end—Continued

	Baseline		Phase 3			Phase 4			Phase 5			Phase 6			Data Collection End	
Domain category	*n*	%	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	*n*	%
Number of postsecondary institutions applied to																
0	659,033	20.37	170,825	15.28	25.59	233,275	15.89	27.47	284,217	16.29	41.86	352,319	17.62	25.40	431,976	18.34
1	1,044,881	32.30	355,085	31.76	34.04	472,397	32.18	33.08	558,337	32.00	33.95	640,410	32.03	33.07	761,735	32.34
2 to 4	1,015,962	31.40	389,446	34.83	27.30	505,638	34.44	26.98	600,234	34.40	16.56	668,626	33.44	28.12	772,301	32.79
5 or more	515,312	15.93	202,622	18.12	13.08	256,681	17.49	12.47	302,156	17.32	7.63	338,319	16.92	13.41	389,617	16.54
Number of high schools attended																
1	2,698,550	83.41	937,619	83.87	81.14	1,229,148	83.73	82.20	1,460,100	83.68	76.61	1,657,898	82.91	83.69	1,952,812	82.90
2	461,858	14.28	153,920	13.77	17.39	207,581	14.14	15.38	246,552	14.13	16.65	292,413	14.62	14.04	345,589	14.67
3 or more	74,780	2.31	26,439	2.36	1.47	31,262	2.13	2.42	38,292	2.19	6.74	49,363	2.47	2.27	57,229	2.43
Apprenticing as of Nov. 1, 2013																
Yes	105,018	3.25	28,123	2.52	4.13	39,220	2.67	3.47	45,831	2.63	7.26	57,096	2.86	4.18	70,588	3.00
No	2,610,097	80.68	929,324	83.13	77.48	1,213,378	82.66	77.89	1,436,935	82.35	69.78	1,629,458	81.49	78.47	1,912,562	81.19
Don't know	520,074	16.08	160,530	14.36	18.39	215,393	14.67	18.64	262,179	15.03	22.97	313,121	15.66	17.36	372,480	15.81

See notes at end of table.

Working for pay as of Nov. 1, 2013																
Yes	1,843,058	56.97	577,427	51.65	61.88	768,829	52.37	68.36	934,010	53.53	70.26	1,097,524	54.89	62.02	1,304,867	55.39
No	985,264	30.45	380,603	34.04	26.87	492,995	33.58	23.30	579,220	33.19	20.02	638,388	31.92	26.70	742,472	31.52
Don't know	406,867	12.58	159,947	14.31	11.25	206,167	14.04	8.34	231,715	13.28	9.72	263,763	13.19	11.28	308,291	13.09
Serving in military as of Nov. 1, 2013																
Yes	127,723	3.95	32,779	2.93	6.40	48,870	3.33	6.33	59,633	3.42	6.91	74,119	3.71	4.05	85,405	3.63
No	2,971,449	91.85	1,040,228	93.05	88.18	1,360,446	92.67	89.28	1,616,655	92.65	83.66	1,841,751	92.10	91.68	2,169,712	92.11
Don't know	136,017	4.20	44,971	4.02	5.42	58,674	4.00	4.39	68,656	3.93	9.43	83,804	4.19	4.27	100,512	4.27

71 **Table C-18. Weighted estimates of bias likelihood model variables and other key variables, at baseline, phase target selection,**
72 **and data collection end—Continued**

	Baseline		Phase 3			Phase 4			Phase 5			Phase 6			Data Collection End	
Domain category	*n*	%	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	*n*	%
Starting family / taking care of children as of Nov. 1, 2013																
Yes	193,540	5.98	45,750	4.09	8.47	66,612	4.54	8.53	86,822	4.98	13.77	110,106	5.51	7.11	134,246	5.70
No	2,929,622	90.55	1,035,030	92.58	88.20	1,354,678	92.28	87.50	1,598,953	91.63	83.40	1,822,092	91.12	88.95	2,140,977	90.89
Don't know	112,027	3.46	37,198	3.33	3.33	46,700	3.18	3.97	59,169	3.39	2.82	67,476	3.37	3.94	80,406	3.41
Completed FAFSA for teenager's education																
Yes	2,189,140	67.67	813,644	72.78	62.58	1,051,658	71.64	61.46	1,242,985	71.23	58.87	1,408,331	70.43	62.35	1,638,479	69.56
No	727,806	22.50	213,710	19.12	25.27	291,031	19.83	25.20	347,886	19.94	27.56	407,374	20.37	26.00	490,154	20.81
Don't know	78,758	2.43	20,122	1.80	3.11	28,405	1.93	4.52	40,397	2.32	1.94	45,370	2.27	2.81	54,897	2.33
Don't know if teenager or another family member completed FAFSA	239,485	7.40	70,502	6.31	9.04	96,896	6.60	8.81	113,676	6.51	11.63	138,599	6.93	8.84	172,099	7.31

See notes at end of table.

Currently working for pay																
Yes	1,610,047	49.77	558,184	49.93	51.80	735,652	50.11	47.91	874,890	50.14	49.02	992,708	49.64	49.01	1,175,024	49.88
No	1,625,142	50.23	559,794	50.07	48.20	732,338	49.89	52.09	870,055	49.86	50.98	1,006,966	50.36	50.99	1,180,606	50.12
Attended CTE center																
No	3,178,886	98.26	1,101,854	98.56	98.05	1,445,388	98.46	98.04	1,717,779	98.44	96.47	1,965,969	98.31	98.06	2,314,937	98.27
Yes	56,302	1.74	16,124	1.44	1.95	22,602	1.54	1.96	27,165	1.56	3.53	33,705	1.69	1.94	40,693	1.73

Table C-18. Weighted estimates of bias likelihood model variables and other key variables, at baseline, phase target selection, and data collection end—Continued

	Baseline		Phase 3			Phase 4			Phase 5			Phase 6			Data Collection End	
Domain category	*n*	%	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	Respondent *n*	Respondent %	Targeted %	*n*	%
English language learner status																
Not English as second language	3,145,642	97.23	1,095,680	98.01	96.58	1,435,266	97.77	96.41	1,705,235	97.72	95.32	1,949,382	97.48	96.50	2,297,090	97.51
English as a second language	89,547	2.77	22,298	1.99	3.42	32,724	2.23	3.59	39,709	2.28	4.68	50,293	2.52	3.50	58,540	2.49

[1] Race categories exclude persons of Hispanic ethnicity.

NOTE: FAFSA = Free Application for Federal Student Aid; CTE = career and technical education.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09) Second Follow-up.

Model effectiveness in targeting underrepresented cases. The bias likelihood model was designed to identify nonrespondent cases most unlike the respondent set at each phase of data collection Therefore, for a model to be successful in identifying underrepresented cases, the distribution within a variable of cases identified for targeting should differ from the respondent set within that variable, particularly if there is an imbalance from the baseline distribution. As an example, consider the model variable *Sex*. At baseline, the total weighted population consisted of approximately 51 percent male and 49 percent female. At the beginning of phase 3 (the start of responsive design case targeting), the weighted set of respondents was 42 percent male and 58 percent female, indicating an imbalance. Therefore, the targeted set of cases should overrepresent males, as indicated by the phase 3 distribution within the targeted set: 70 percent male and 30 percent female. Many of the model variables listed in table F-18 demonstrate this pattern, suggesting that the bias likelihood model was effective in identifying cases underrepresented on those key variables included in the model.

Model effectiveness in reducing sample imbalance within key survey variables. If the bias likelihood model was effective in targeting underrepresented cases and the interventions were effective, the expectation is to observe a reduction in imbalance, over time, as a result of increasing response among targeted cases. As an example, consider the model variable, *Taking postsecondary classes as of November 1, 2013* (see table F-18). At baseline, 67 percent of the overall sample was taking postsecondary classes while 21, and 12 percent were not or did not know, respectively. The respondent set at the start of phase 3 was 76 percent taking postsecondary classes, while 15 and 9 percent were not and did not know, respectively. Sample imbalance at phase 3 was clearly present with overrepresentation among those taking postsecondary classes. Over the subsequent data collection phases, the percentage of the respondent set taking postsecondary classes decreased (76 to 75 to 74 to 72 percent at the start of phases 3, 4, 5, and 6 and ending at 70 percent at the close of data collection) while the set of those not taking postsecondary classes increased (from 15 to 16 to 17 to 18 percent at the start of phases 3, 4, 5, and 6, and ending at 19 percent the conclusion of data collection). This pattern brought the variable distribution closer to the baseline distribution, addressing some of

679 the imbalance present at the start of phase 3. Changes in this survey
680 estimate between the start of phase 3 and the end of data collection
681 appear to move in the direction of the estimates for the entire
682 sample. The pattern observed in this example is illustrative of the
683 general trend evident across many of the model survey variables.

684