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

96   identified, these critical nonrespondents could be targeted for
97   tailored incentives dependent on their respective subgroup.

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

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

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

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

122 To determine optimal incentive amounts, a calibration
123 subsample was selected from each of the aforementioned
124 subgroups to begin data collection ahead of the main sample.
125 The experimental sample was treated in advance of the
126 remaining cases. Results from the calibration sample

3   [1] In the spirit of a responsive design, the set of cases to be treated as "ultra-
4   cooperative" was expanded for the main sample (i.e., cases not in the
5   calibration sample) with the goal of maximizing the efficient use of project
6   resources because response rates were reasonably high. See section 4.2.1.6
7   for further details and for the expanded definition. The definition provided
8   above corresponds to that used for sample members in the calibration
9   sample.

127   experiments were used to determine the incentive levels – a
128   *baseline* incentive and two subsequent incentive increases, or
129   *boosts* – offered to the remaining (i.e., noncalibration) sample in
130   each of the three subgroups.

131   The data collection design for the second follow-up included a
132   responsive design with multiple intervention phases. These
133   phases included specific protocols for handling each of the three
134   subgroups of sample members to reduce the potential for biased
135   survey estimates or reduce data collection costs (Peytchev
136   2013). For more details on the second follow-up data collection
137   design, see section 4.2.1.

C.2 Responsive Design Model Development

138

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

C.2.1 Response Likelihood Model Development

150

151   The response likelihood model was developed using data from
152   earlier rounds, and was designed to predict the a priori likelihood
153   of a case becoming a respondent. The response likelihood model
154   allowed the data collection team to identify cases with a low
155   probability of responding and avoid applying relatively expensive
156   interventions, such as field interviewing, to these cases. To make
157   the interventions more cost efficient, the primary objective of the
158   response likelihood model was to inform decisions about the
159   exclusion of cases that were identified for targeting based on the
160   bias likelihood model but which had extremely low likelihood of
161   participation. From a model-building perspective, the objective
162   was to maximize prediction of participation, regardless of any

163   association between the predictor variables and the HSLS:09
164   survey variables.

165   From prior analysis in the base year, first follow-up, and 2013
166   Update, candidate variables known to be predictive of response
167   behavior (i.e., prior-round response outcomes) were considered
168   for the response likelihood model. To determine which covariates
169   to include in the model, stepwise logistic regression was run with
170   the model entry criteria set to p = .5—meaning that any
171   predictor variable with an initial probability value of .5 or less
172   was included in the stepwise regressions—and model retention
173   criteria set to p = .1—meaning that any variable with a
174   probability value of .1 or less was retained in the final model. The
175   result of this approach is the retention of a set of covariates
176   capable of predicting a case's likelihood of becoming a
177   respondent. Table F-1 lists all predictor variables considered for
178   inclusion in the response likelihood model and their final
179   inclusion disposition (i.e., which variables were retained and
180   which were released from the final model).

181
182

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

183   [1] Race categories exclude persons of Hispanic ethnicity.

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

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

Data source	Variable	Odds ratio	Lower bound	Upper bound	Interpretation
			95% confidence interval		
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

Data source	Variable	Odds ratio	Lower bound	Upper bound	Interpretation
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

192   See notes at end of table.

193 **Table C-2. Odds ratios and confidence intervals for variables in the response likelihood model:**
194 **   2016—Continued**

Data source	Variable	Odds ratio	Lower bound	Upper bound	Interpretation
			95% confidence interval		
	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

195 NOTE: Race categories exclude persons of Hispanic ethnicity.

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

198   **Response likelihood model definition.** Using the final
199   covariates selected (primarily paradata variables), a model was
200   developed to predict the response outcome in the 2013 Update,
201   the last data collection round prior to the second follow-up. The
202   response likelihood model used a logit function to generate, for
203   each case, a continuous probability of response (bounded by 0
204   and 1), called a response likelihood score, in which a value of 1
205   indicated a case was predicted to respond and 0 indicated a case

206   was predicted not to respond. Response likelihood values were
207   calculated one time prior to the beginning of data collection.

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

215
$$\Pr(y_i^{2013\ Update}=1)=\text{logit}^{-1}(\beta_0+\beta_1\bullet\text{female}_i+\beta_2\bullet\text{base year response}_i+\ldots)$$

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

218
$$\hat{p}_i^{response}=\frac{e^{(\beta_0+\beta_1\bullet\text{female}_i+\beta_2\bullet\text{base year response}_i+\ldots)}}{1+e^{(\beta_0+\beta_1\bullet\text{female}_i+\beta_2\bullet\text{base year response}_i+\ldots)}}=\Pr(y_i^{2013\ Update}=1)$$

219   **Overall response likelihood distribution.** Across the entire
220   second follow-up fielded sample (n = 23,316)[2], the overall mean
221   response likelihood score was .80. As indicated by this mean,
222   many sample members were clustered at the upper end of the
223   distribution. Within the three subgroups of interest, subgroup A
224   (HSNC; n = 2,545) had a mean response probability of .65. As
225   expected, these cases were found to have the lowest average
226   response likelihood value among all of the subgroups.
227   Conversely, subgroup B (UC; n = 4,144) had a mean response
228   probability of .96, indicating that these cases were highly likely
229   to be respondents per the response likelihood model. Subgroup C
230   cases (HS other; n = 16,627) had a mean response probability
231   of .78, very close to the fielded sample's overall mean.

232   As noted in section 4.2.1.2, the model-derived response
233   likelihood scores were used to assist in determining intervention
234   resource allocation only in phases 5 and 6 to avoid pursuing
235   cases in field interviewing that were unlikely to respond. Section
236   4.2.1.4 provides further details on the use of these scores.

18   [2] See section 2.4 for a description of the second follow-up sample design.

237   ## C.2.2 Bias Likelihood Model Development

238   The goal of the bias likelihood model was to identify cases most
239   likely to contribute to nonresponse bias because their
240   characteristics were underrepresented among the set of
241   respondents. This approach provided an overview of where
242   sample underrepresentation might be occurring in the
243   respondent set. To achieve this goal, the criteria for inclusion of
244   variables in the bias likelihood model differed from the criteria
245   for inclusion in the response likelihood model. Maximizing the
246   prediction of survey participation was not the main objective. In
247   the bias likelihood model, variables of high analytic value were
248   sought for inclusion in the model. Therefore, model fit and
249   statistical significance were not primary determining factors in
250   deciding which variables to include in the bias likelihood model.
251   Rather, variables were selected for inclusion in the bias likelihood
252   model principally due to their analytic importance to the study.
253   Conversely, variables that were highly predictive of participation
254   but not necessarily associated with the survey variables, such as
255   paradata on the ease of obtaining participation on the previous
256   administration, were excluded as they could have a
257   disproportionate influence on the predicted propensities without
258   contributing additional information on bias in the second follow-
259   up. Once the set of key variables was identified, stepwise logistic
260   regression was used to help improve overall model fit. Bias
261   likelihood model variables, and their corresponding level of data
262   requiring imputation, are presented in table F-3. Note that many
263   key survey variables from prior rounds contained missing values
264   which required imputation to be included in the bias likelihood
265   model. Further discussion of the imputation process follows in
266   the text below.

267   **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

268   See notes at end of table.

269   **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

270   [1] Race categories exclude persons of Hispanic ethnicity.

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

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

275   **Imputation process.** Assessment of balance between
276   respondents and nonrespondents required having nonmissing
277   data for both groups. To be used as bias likelihood model
278   covariates, many key survey variables containing missing values
279   required imputation. Missing data were imputed for these survey
280   variables using stochastic imputation. Prior-round
281   nonrespondents were included in imputation since the goal was
282   to achieve a complete dataset for all second follow-up sample
283   members. Specifically, a weighted sequential hot-deck (WSHD)
284   statistical imputation procedure (Cox 1980; Iannacchione 1982),

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

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

---

23   [3] The student base weight was used as it is nonmissing for all sample
24   members. For further details on weights available in the second follow-up,
25   including the student base weight, see chapter 6.
26   [4] A serpentine sort is a sorting method in which records are ordered in an
27   alternating ascending and descending pattern, thereby causing any two
28   consecutive records in the sorted file to have similar values for the sort
29   variables.

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

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

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

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

342

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

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

346

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.

367   **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

368   [1] Beginning with phase 5, calibration sample and main sample cases were combined for data collection treatments.
369   NOTE: Subgroup A = high school late/alternative/noncompleters; subgroup B = ultra-cooperative respondents; subgroup C = all
370   other high school completers and unknown cases; CATI = computer-administered telephone interviewing.
371   SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009
372   (HSLS:09) Second Follow-up.

373   **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	

374   [1] Subgroup B (ultra-cooperative respondents) cases offered a nonzero baseline incentive (i.e., $30, $40, or $50) were not eligible to
375   be targeted to receive subsequent treatments (i.e., incentive boost 1 or boost 2).
376   SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009
377   (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

---

[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.

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

414   **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

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

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

439   subgroup B, no baseline incentive was offered to subgroup B
440   cases in the main sample.

441 **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

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

445 **Table C-9.   Comparison of subgroup B response rates with response rates from selected**
446 **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

447 [1] Excludes partially completed cases.
448 [2] The B&B:08/08 and the B&B:08/12 early response phases consisted of the first 4 weeks of data collection.
449 NOTE: HSLS:09 = High School Longitudinal Study of 2009; BPS:12/14 = 2012/14 Beginning Postsecondary Students Longitudinal
450 Study; B&B:08/12 = 2008/12 Baccalaureate and Beyond Longitudinal Study; BPS:08/09 = 2008/2009 Beginning Postsecondary
451 Students Longitudinal Study; NPSAS:08 = 2007–08 National Postsecondary Student Aid Study.
452 SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009
453 (HSLS:09) Second Follow-up; U.S. Department of Education, National Center for Education Statistics, 2012/14 Beginning
454 Postsecondary Students Longitudinal Study (BPS:12/14); U.S. Department of Education, National Center for Education Statistics,
455 2008/12 Baccalaureate and Beyond Longitudinal Study (B&B:08/12).

456   **Subgroup C (HS other).** Table F-10 provides subgroup C (HS
457   other) response rates by baseline incentive level. Within
458   subgroup C, the highest response rate, 43 percent, was observed
459   among cases assigned a $30 incentive. No significant difference
460   was detected between the response rate associated with the $30
461   baseline incentive and that of either the $35 incentive or $40
462   incentive. Response rates among cases assigned the $30
463   incentive were significantly higher than those for $15 and $20
464   ($\chi2$ (1, N = 658) = 17.28, p < .05 and $\chi2$ (1, N = 658) = 6.59, p
465   < .05, respectively).

466   No significant difference was detected at the .05 level between
467   comparisons of response rates for cases assigned $30 (43
468   percent) and $25 (37 percent) ($\chi2$ (1, N = 658) = 2.53, p = .11).
469   Given that subgroup C constitutes the largest subgroup in the
470   main sample, with more than 14,000 sample members, a 6
471   percent difference in response rate would result in a nontrivial
472   difference in yield; as such, a baseline incentive of $30 was
473   offered to all subgroup C main sample cases.

474   **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

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

478   ### C.3.2 *Phase 3 (Incentive Boost 1 Offer)*

479   Phase 3 of the calibration study introduced an incentive boost
480   that was offered to a subset of pending nonrespondents in
481   addition to the baseline amount offered in the prior phases. The
482   bias likelihood model was deployed prior to the start of phase 3
483   and was used to target subgroup B and subgroup C cases to
484   receive an incentive boost (boost 1) in addition to their baseline
485   incentive, should they complete the survey. Given the relative

486   importance of obtaining responses from subgroup A cases, all
487   remaining nonrespondent cases in subgroup A were targeted for
488   an incentive boost offer.

489   ***Subgroup A (HSNC).*** Table F-11 displays subgroup A response
490   rates during phase 3 by incentive boost level and baseline
491   incentive level. For subgroup A cases that received no baseline
492   incentive, no significant difference was detected between the
493   response rates of sample members who were offered the $15 (10
494   percent) and $25 (15 percent) boost 1 incentive. No significant
495   differences were detected between the response rates of sample
496   members who were offered the $15 (17 percent) and $25 (12
497   percent) boost 1 incentive, when the baseline incentive was $30.
498   Additionally, there was no significant difference detected
499   between the response rates of sample members who were
500   offered the $15 (12 percent) and $25 (19 percent) boost 1
501   incentive, when the baseline incentive was $40. Lastly, no
502   significant differences were detected between the response rates
503   of sample members who were offered the $15 (12 percent) and
504    $25 (17 percent) boost 1 incentive, when the baseline incentive
505   was $50. Given that no significant differences were found
506   between the $15 and $25 boost incentives, based on the results
507   available on June 7, 2016, a boost 1 incentive of $15 was offered
508   to all phase 3 cases in the subgroup A main sample.

509   **Table C-11. Subgroup A response rates in phase 3, by boost 1 incentive amount as of**
510   **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

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

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

526   **Table C-12. Subgroup B response rates in phase 3, by boost 1 incentive amount as of**
527   **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

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

531   ***Subgroup C (HS other).*** Table C-13 displays subgroup C
532   response rates during phase 3 by incentive level, among the 661
533   cases selected for an incentive boost offer based on the bias
534   likelihood model. No significant difference was detected between
535   the phase 3 response rates of sample members offered $10
536   (13.9 percent) and $20 (15.5 percent) boost 1 incentives,
537   regardless of the baseline incentive offered. As such, a boost 1
538   incentive of $10 was offered to all phase 3 targeted cases in the
539   subgroup C main sample.

540   **Table C-13. Subgroup C response rates in phase 3, by boost 1 incentive amount as of**
541   **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

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

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

546   Phase 4 of the calibration study introduced a second incentive
547   boost that was offered to a subset of pending nonrespondents in
548   addition to the baseline amount and first boost, as applicable.
549   The bias likelihood model was deployed again prior to the start of
550   phase 4 and was again used to identify cases in subgroup B and
551   subgroup C for targeted interventions (i.e., to receive an
552   incentive boost offer). Note that cases were selected for the
553   boost 2 offer independently from the selection of cases for boost
554   1. A case targeted for a boost 1 incentive offer might or might
555   not be selected to receive a boost 2 incentive offer depending on
556   how its bias likelihood score shifted between the phases. As was
557   done in phase 3, all remaining nonrespondent cases in subgroup
558   A were targeted for an incentive boost 2 offer. An initial analysis
559   of the boost 2 incentive was conducted after 4 weeks (July 15,
560   2016) to determine the optimal incentive amount for the main
561   sample. However, a second analysis after approximately 11
562   weeks (September 7, 2016) revealed that the results had shifted
563   for subgroups A and C, as detailed below.

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

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

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

587   ***Incentive boost 2b.*** While response rates for cases assigned to
588   $10 and $20 boost 2 incentive levels were statistically equivalent
589   (i.e., no significant differences were detected) at 4 weeks for
590   each of the subgroups, when reassessed after about 11 weeks
591   (September 7, 2016) the differences between cases assigned
592   $10 and $20 had become large and statistically significant for
593   subgroup A (χ^2 (1, N = 310) = 6.38, p < .05) and subgroup C (χ^2
594   (1, N = 576) = 4.02, p < .05). (Subgroup B had very small
595   numbers and no detectable difference.) The additional time for
596   the calibration sample cases in phase 4 revealed an effect that
597   was not evident at the end of the first 4 weeks of phase 4. In the
598   intervening weeks, staff increased locating, prompting, and case
599   review efforts for all pending cases (regardless of incentive
600   amount assignment). Results after 4 weeks in phase 4 and after

601   11 weeks in phase 4 are presented below in tables F-14, F-15,
602   and F-16.

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

Boost 2 incentive offer	Sample members (*n*)	Boost 2 results after 4 weeks		Boost 2 results after 11 weeks	
		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

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

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

Boost 2 incentive offer	Sample members (*n*)	Boost 2 results after 4 weeks		Boost 2 results after 11 weeks	
		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

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

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

Boost 2 incentive offer	Sample members (*n*)	Boost 2 results after 4 weeks		Boost 2 results after 11 weeks	
		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

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

618   Based on results after 11 weeks in phase 4, an adaptive
619   component was added to the responsive design protocol in which
620   an additional boost (incentive boost 2b) of $10 was offered to
621   subgroup A main sample nonrespondents and subgroup C main
622   sample boost 2-targeted cases; no additional boost was offered
623   to subgroup B cases.

C.4   Assessment of Responsive Design Models

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

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

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

643   **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

644   [1] Note the total sample (23,316) represents to total fielded sample and excludes sample members that withdrew from the study
645   between the end of the 2013 Update collection and the beginning of the second follow-up data collection or were found to be
646   deceased.
647   SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009
648   (HSLS:09) Second Follow-up.

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

655   ### C.4.2   Assessment of Bias Likelihood Model on Sample Representativeness

656   As described in section 4.2.1.3, the bias likelihood model was
657   used to identify cases that were most unlike the set of sample
658   members that had responded at each time-point the model was
659   fit. The model used key survey and frame variables as model
660   covariates with current nonresponse (as of each model run) as
661   the dependent variable to identify nonrespondents most likely to

662   contribute to bias in key survey variables unless converted to
663   respondents. The bias likelihood model was fit at the beginning
664   of phases 3 and 4 for the calibration and main samples (i.e., prior
665   to both boost interventions) and at the beginning of phases 5
666   and 6[6] for the combined sample.

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

677

---

58   [6] Beginning with phase 5, calibration sample and main sample cases were
59   combined for data collection treatments. Note that phases 5 and 6 were not
60   part of the calibration experiment, and are therefore not covered in this
61   appendix. For details on these phases, see section 4.2.1.4.

678
679

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

Domain category	Baseline n	Baseline %	Phase 3 Respondent n	Phase 3 Respondent %	Phase 3 Targeted %	Phase 4 Respondent n	Phase 4 Respondent %	Phase 4 Targeted %	Phase 5 Respondent n	Phase 5 Respondent %	Phase 5 Targeted %	Phase 6 Respondent n	Phase 6 Respondent %	Phase 6 Targeted %	Data Collection End n	Data Collection End %
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

62
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 n	Baseline %	Phase 3 Respondent n	Phase 3 Respondent %	Phase 3 Targeted %	Phase 4 Respondent n	Phase 4 Respondent %	Phase 4 Targeted %	Phase 5 Respondent n	Phase 5 Respondent %	Phase 5 Targeted %	Phase 6 Respondent n	Phase 6 Respondent %	Phase 6 Targeted %	Data Collection End n	Data Collection End %
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.

Domain category	Baseline n	Baseline %	Phase 3 Respondent n	Phase 3 Respondent %	Phase 3 Targeted %	Phase 4 Respondent n	Phase 4 Respondent %	Phase 4 Targeted %	Phase 5 Respondent n	Phase 5 Respondent %	Phase 5 Targeted %	Phase 6 Respondent n	Phase 6 Respondent %	Phase 6 Targeted %	Data Collection End n	Data Collection End %
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

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 n	Baseline %	Phase 3 Respondent n	Phase 3 Respondent %	Phase 3 Targeted %	Phase 4 Respondent n	Phase 4 Respondent %	Phase 4 Targeted %	Phase 5 Respondent n	Phase 5 Respondent %	Phase 5 Targeted %	Phase 6 Respondent n	Phase 6 Respondent %	Phase 6 Targeted %	Data Collection End n	Data Collection End %
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.

Domain category	Baseline n	Baseline %	Phase 3 Respondent n	Phase 3 Respondent %	Phase 3 Targeted %	Phase 4 Respondent n	Phase 4 Respondent %	Phase 4 Targeted %	Phase 5 Respondent n	Phase 5 Respondent %	Phase 5 Targeted %	Phase 6 Respondent n	Phase 6 Respondent %	Phase 6 Targeted %	Data Collection End n	Data Collection End %
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

66
67

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 n	Baseline %	Phase 3 Respondent n	Phase 3 Respondent %	Phase 3 Targeted %	Phase 4 Respondent n	Phase 4 Respondent %	Phase 4 Targeted %	Phase 5 Respondent n	Phase 5 Respondent %	Phase 5 Targeted %	Phase 6 Respondent n	Phase 6 Respondent %	Phase 6 Targeted %	Data Collection End n	Data Collection End %
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.

Domain category	Baseline n	Baseline %	Phase 3 Respondent n	Phase 3 Respondent %	Phase 3 Targeted %	Phase 4 Respondent n	Phase 4 Respondent %	Phase 4 Targeted %	Phase 5 Respondent n	Phase 5 Respondent %	Phase 5 Targeted %	Phase 6 Respondent n	Phase 6 Respondent %	Phase 6 Targeted %	Data Collection End n	Data Collection End %
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

68
69 **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 n	Baseline %	Phase 3 Respondent n	Phase 3 Respondent %	Phase 3 Targeted %	Phase 4 Respondent n	Phase 4 Respondent %	Phase 4 Targeted %	Phase 5 Respondent n	Phase 5 Respondent %	Phase 5 Targeted %	Phase 6 Respondent n	Phase 6 Respondent %	Phase 6 Targeted %	Data Collection End n	Data Collection End %
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.

Domain category	Baseline n	Baseline %	Phase 3 Respondent n	Phase 3 Respondent %	Phase 3 Targeted %	Phase 4 Respondent n	Phase 4 Respondent %	Phase 4 Targeted %	Phase 5 Respondent n	Phase 5 Respondent %	Phase 5 Targeted %	Phase 6 Respondent n	Phase 6 Respondent %	Phase 6 Targeted %	Data Collection End n	Data Collection End %
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

70
71

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	%
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

See notes at end of table.

72
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**

Domain category	Baseline n	Baseline %	Phase 3 Respondent n	Phase 3 Respondent %	Phase 3 Targeted %	Phase 4 Respondent n	Phase 4 Respondent %	Phase 4 Targeted %	Phase 5 Respondent n	Phase 5 Respondent %	Phase 5 Targeted %	Phase 6 Respondent n	Phase 6 Respondent %	Phase 6 Targeted %	Data Collection End n	Data Collection End %
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

74
75 **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 n	Baseline %	Phase 3 Respondent n	Phase 3 Respondent %	Phase 3 Targeted %	Phase 4 Respondent n	Phase 4 Respondent %	Phase 4 Targeted %	Phase 5 Respondent n	Phase 5 Respondent %	Phase 5 Targeted %	Phase 6 Respondent n	Phase 6 Respondent %	Phase 6 Targeted %	Data Collection End n	Data Collection End %
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
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.

Domain category	Baseline n	Baseline %	Phase 3 Respondent n	Phase 3 Respondent %	Phase 3 Targeted %	Phase 4 Respondent n	Phase 4 Respondent %	Phase 4 Targeted %	Phase 5 Respondent n	Phase 5 Respondent %	Phase 5 Targeted %	Phase 6 Respondent n	Phase 6 Respondent %	Phase 6 Targeted %	Data Collection End n	Data Collection End %
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

76
77 **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 n	Baseline %	Phase 3 Respondent n	Phase 3 Respondent %	Phase 3 Targeted %	Phase 4 Respondent n	Phase 4 Respondent %	Phase 4 Targeted %	Phase 5 Respondent n	Phase 5 Respondent %	Phase 5 Targeted %	Phase 6 Respondent n	Phase 6 Respondent %	Phase 6 Targeted %	Data Collection End n	Data Collection End %
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
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.

78
79 **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 n	Baseline %	Phase 3 Respondent n	Phase 3 Respondent %	Phase 3 Targeted %	Phase 4 Respondent n	Phase 4 Respondent %	Phase 4 Targeted %	Phase 5 Respondent n	Phase 5 Respondent %	Phase 5 Targeted %	Phase 6 Respondent n	Phase 6 Respondent %	Phase 6 Targeted %	Data Collection End n	Data Collection End %
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
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

80
81
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 n	Baseline %	Phase 3 Respondent n	Phase 3 Respondent %	Phase 3 Targeted %	Phase 4 Respondent n	Phase 4 Respondent %	Phase 4 Targeted %	Phase 5 Respondent n	Phase 5 Respondent %	Phase 5 Targeted %	Phase 6 Respondent n	Phase 6 Respondent %	Phase 6 Targeted %	Data Collection End n	Data Collection End %
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.

Domain category	Baseline n	Baseline %	Phase 3 Respondent n	Phase 3 Respondent %	Phase 3 Targeted %	Phase 4 Respondent n	Phase 4 Respondent %	Phase 4 Targeted %	Phase 5 Respondent n	Phase 5 Respondent %	Phase 5 Targeted %	Phase 6 Respondent n	Phase 6 Respondent %	Phase 6 Targeted %	Data Collection End n	Data Collection End %
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

82
83

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 n	Baseline %	Phase 3 Respondent n	Phase 3 Respondent %	Phase 3 Targeted %	Phase 4 Respondent n	Phase 4 Respondent %	Phase 4 Targeted %	Phase 5 Respondent n	Phase 5 Respondent %	Phase 5 Targeted %	Phase 6 Respondent n	Phase 6 Respondent %	Phase 6 Targeted %	Data Collection End n	Data Collection End %
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.

Domain category	Baseline n	Baseline %	Phase 3 Respondent n	Phase 3 Respondent %	Phase 3 Targeted %	Phase 4 Respondent n	Phase 4 Respondent %	Phase 4 Targeted %	Phase 5 Respondent n	Phase 5 Respondent %	Phase 5 Targeted %	Phase 6 Respondent n	Phase 6 Respondent %	Phase 6 Targeted %	Data Collection End n	Data Collection End %
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

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 n	Baseline %	Phase 3 Respondent n	Phase 3 Respondent %	Phase 3 Targeted %	Phase 4 Respondent n	Phase 4 Respondent %	Phase 4 Targeted %	Phase 5 Respondent n	Phase 5 Respondent %	Phase 5 Targeted %	Phase 6 Respondent n	Phase 6 Respondent %	Phase 6 Targeted %	Data Collection End n	Data Collection End %
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
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.

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

704   ***Model effectiveness in reducing sample imbalance within***
705   ***key survey variables.*** If the bias likelihood model was effective
706   in targeting underrepresented cases and the interventions were
707   effective, the expectation is to observe a reduction in imbalance,
708   over time, as a result of increasing response among targeted
709   cases. As an example, consider the model variable, *Taking*
710   *postsecondary classes as of November 1, 2013* (see table F-18).
711   At baseline, 67 percent of the overall sample was taking
712   postsecondary classes while 21, and 12 percent were not or did
713   not know, respectively. The respondent set at the start of phase
714   3 was 76 percent taking postsecondary classes, while 15 and 9
715   percent were not and did not know, respectively. Sample
716   imbalance at phase 3 was clearly present with
717   overrepresentation among those taking postsecondary classes.
718   Over the subsequent data collection phases, the percentage of
719   the respondent set taking postsecondary classes decreased (76
720   to 75 to 74 to 72 percent at the start of phases 3, 4, 5, and 6 and
721   ending at 70 percent at the close of data collection) while the set
722   of those not taking postsecondary classes increased (from 15 to
723   16 to 17 to 18 percent at the start of phases 3, 4, 5, and 6, and
724   ending at 19 percent the conclusion of data collection). This

725   pattern brought the variable distribution closer to the baseline
726   distribution, addressing some of the imbalance present at the
727   start of phase 3. Changes in this survey estimate between the
728   start of phase 3 and the end of data collection appear to move in
729   the direction of the estimates for the entire sample. The pattern
730   observed in this example is illustrative of the general trend
731   evident across many of the model survey variables.

732