

September 22, 2011

High School Longitudinal Study of
2009 (HSL:09)
College Update and Transcript Field
Test (2012)

Supporting Statement
Part B

Request for OMB Review
OMB# 1850-0852 v.10

Submitted by

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

TABLE OF CONTENTS

Section	Page
B. Collection of Information Employing Statistical Methods.....	1
B.1 Target Universe and Sampling Frames.....	1
B.2 Statistical Procedures for Collecting Information.....	1
B.2.a School Sample.....	1
B.2.b Student Sample.....	1
B.2.c Weighting.....	2
B.2.d Imputation of Missing Data.....	4
B.2.e Variance Estimation.....	4
B.3 Methods for Maximizing Response Rates.....	5
B.4 Study Contacts.....	9

EXHIBITS

Number	Page
Exhibit B-1. Consultants on Statistical Aspects of HSLs:09.....	9

B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

This section describes the target universe for this study and the sampling and statistical methodologies proposed for the HSLS:09 College Update and transcript field test and main study collections. Part B also addresses suggested methods for maximizing response rates and for testing procedures and methods and introduces the technical staff responsible for design and administration of the study.

B.1 Target Universe and Sampling Frames

The base-year target populations for HSLS:09 consisted of (1) public and private schools within the U.S. providing instruction to 9th- and 11th-grade students, and (2) the 9th graders attending these schools in the fall semester of 2009 (main study) or 2008 (field test). As with the first follow-up studies, the target populations for the HSLS:09 College Update and transcript studies are the same as specified in the base year. Field test studies themselves are not designed to make target population estimates; however the field test samples are taken from the target populations to test all of the study protocols and procedures.

B.2 Statistical Procedures for Collecting Information

B.2.a School Sample

In 2010, consent to participate was obtained from 24 school administrators across five states for the first follow-up field test. These same 24 schools will be contacted for the HSLS:09 transcript field test in the fall of 2012. Transcripts will be requested from any additional schools that the sampled students have attended since the 9th grade.

For the main study, all 944 base-year participating high schools will be contacted for transcript collection in the fall of 2013. Additional schools attended by sample members that are identified during the first follow-up and College Update will also be included in the transcript collection effort.

B.2.b Student Sample

A total of 754 out of 827 students (91.2 percent) participated in either the base-year or first follow-up field tests. These 754 respondents will be contacted to participate in the HSLS:09 College Update field test. Students who did not participate in the base-year and first follow-up data collections will be excluded from the College Update and transcript data collections. However, questionnaire incapable students from the base year and first follow-up will be included. Students

who responded to either the base year or first follow-up or both will be included in the College Update and transcript efforts. One knowledgeable parent will be recruited to provide proxy information if the student is unable or unwilling to participate in the College Update field test. The student's high school records will be collected, keyed, and coded as part of the HSLs:09 transcript field test.

The same procedures will be implemented for the College Update and transcript main studies. After finalizing the response status from the first follow-up main study conducted in Spring 2012, all students who participated in either the base-year or first follow-up main studies (i.e. eligible respondents and questionnaire-incapable students) will be included in the sample for the HSLs:09 College Update main study. Student information provided either directly by the student or indirectly by a knowledgeable parent will be used to minimize unit nonresponse on the College Update survey.

B.2.c Weighting

Analysis weights along with survey data are used to produce population estimates. The weights reflect the inclusion probabilities for the sampled units (i.e., base weights generated in the base year study) and adjustments to lower (1) unit nonresponse bias, (2) undercoverage bias, and (3) the variability of the resulting weights. Analysis weights will be produced only for the HSLs:09 College Update and transcript main studies. Population estimation is not the goal of the field test so that no analysis weights are required.

The HSLs:09 longitudinal, multistage design introduces significant complexity to the task of weighting. Two sets of longitudinal weights are anticipated for the analysis of the cumulative HSLs:09 data: one set to reflect response to either the base-year or first follow-up rounds *and* the College Update; and one set to reflect response to either the base-year or first follow-up rounds *and* receipt of high school transcript information.

The HSLs:09 weighting process includes four major steps. Using the base weights created during the HSLs:09 base-year study, an adjustment will be applied for nonresponse to the base-year *and* first follow-up main studies to account for those excluded from the College Update and transcript studies. In the second step, base weights will be adjusted for nonresponse in the current study (i.e., student/parent nonresponse in the College Update main study). The third step will include a calibration adjustment to the sum of the base-year analysis weights to ensure coverage of

the 9th-grade target population. Finally, weights constructed after each adjustment will go through an extensive series of quality control (QC) checks to prevent any computational or procedural errors and to detect extreme outliers that can decrease the precision in the population estimates. These include review of program logs, verification of weight sums before and after adjustments are applied to the weights, and verification of the final weight sums against weight sums from the HSLS:09 first follow-up. Design effects for a set of important survey estimates will be calculated and reviewed for extreme values, thus creating an iterative process until the final set of efficient weights is produced.

In addition to analyzing design effects, unit bias analyses will be conducted to determine whether additional variables not already included in the nonresponse models should be investigated. Statistical tests will be conducted on a variety of student questionnaire items. If non-negligible levels of bias remain, the nonresponse and calibration adjustments will be revisited with the goal of lowering the bias. To estimate bias for a generic population parameter θ , we will calculate the following quantity for a set of variables known for both respondents and nonrespondents:

$$\hat{B}(\hat{\theta}_R) = \hat{\theta}_R - \hat{\theta},$$

where $\hat{\theta}_R$ is the estimated parameter using only the respondent data, and $\hat{\theta} = (1 - \hat{\eta}) \hat{\theta}_R + \hat{\eta} \hat{\theta}_{NR}$, the estimated parameter using both the respondent ($\hat{\theta}_R$) and nonrespondent ($\hat{\theta}_{NR}$) data and the weighted nonresponse rate $\hat{\eta}$. Candidate variables known for all sample cases include those from the original sampling frame as well as survey data collected in a previous wave of HSLS:09.

All HSLS:09 weight adjustments—including nonresponse and calibration—will be calculated with a design-based model using the WTADJUST procedure in SUDAAN®, statistical software with built-in controls on extreme values. Model variables will be identified as being associated with a set of key analysis variables as well as the differential pattern of unit nonresponse. Classification procedures such as regression tree analysis will be used to identify these variables from a candidate list that includes stratification variables and data collected from previous waves of HSLS:09.

B.2.d Imputation of Missing Data

Missing values due to item nonresponse will be imputed after the data are edited for only the main studies to lower item nonresponse bias. Item nonresponse will be measured in the field test studies but will remain unchanged.

Imputation in the College Update main study will be performed for items commonly used to define analysis domains, items that are frequently used in cross-tabulations, and items needed for weighting. Categorical HSLs:09 items that are subject to imputation will be imputed using logical imputation¹ where applicable, followed by a weighted sequential hot deck procedure.² By incorporating the sampling weights, this method of imputation takes into account the unequal probabilities of selection in the original sample while controlling the expected number of times a particular respondent's answer will be used as a donor. Variables used to form the imputation classes will be chosen from a candidate set as being statistically associated with the specific pattern of item nonresponse as well as with the variable being imputed using CHAID and other statistical tests. Candidate variables will be drawn from the current survey as well as information collected in the previous waves of HSLs:09. Consistency of the imputed values will be verified within and across the waves of HSLs:09.

B.2.e Variance Estimation

For the main studies, for variance estimation, sets of 200 balanced repeated replication (BRR) weights will be created for the College Update and transcript samples matching the number of replicates used for the HSLs:09 base year. The BRR weighting process will replicate the procedures used to generate the full sample weight and will follow the same steps successfully implemented on a number of studies, including Education Longitudinal Study of 2002 (ELS:2002)³, the National Study of Postsecondary Faculty (NSOPF)⁴, and previous rounds of HSLs:09⁵. In addition, analysis strata and primary sampling units (PSUs) created from the sampling PSUs will be included on the electronic code book for analysts wanting to use Taylor series variance estimation rather than BRR weights.

¹ One commonly used example of logical imputation is assigning gender based on name.

² Iannacchione, V.G. (1982). "Weighted Sequential Hot Deck Imputation Macros." *In Proceedings of the Seventh Annual SAS User's Group International Conference* (pp.759–763). Cary, NC: SAS Institute, Inc.

³ Additional information on ELS:2002 is located at <http://nces.ed.gov/surveys/els2002/>.

⁴ Additional information on NSOPF is located at <http://nces.ed.gov/surveys/nsopf/>.

⁵ Additional information on HSLs:09 is located at <http://nces.ed.gov/surveys/hsls09/>.

B.3 Methods for Maximizing Response Rates

Procedures for maximizing response rates at the institution and respondent levels are based on successful experiences with prior rounds of HSLs:09 and with other similar studies. In this section, methods for maximizing response rates for the College Update (CU) interview and the high school transcript collection are discussed.

College Update. While information specifically relates to the student sample member, the College Update questionnaire can be completed by either the student or the parent. Allowing the parent or the student to provide information will result in a higher response rate than would otherwise be achieved with a student-only response restriction. Students and parents will have the option to complete the interview online or on the telephone. However, as an evaluation of inter-respondent (student vs. parent) agreement, as part of the re-interview protocol, a subsample of both the parent and their son or daughter will be interviewed and their responses compared.

Materials will be mailed to both students and parents. Since the vast majority of students will be at least 18 years of age at the time of the College Update data collection, separate materials will be sent directly to students and to parents. For students who are not yet 18 years old, parent permission must be received before contacting the student or allowing the student to participate. As is the protocol for the First Follow-up data collection, the parent mailing will include a sealed student letter when parent permission is required. Parents will also be allowed to provide their permission online or during an outbound computer assisted telephone interview (CATI) call. Once parent permission is obtained, subsequent reminder mailings will be sent directly to the student.

Students and parents will receive a description of the study, a note stating the importance of the College Update, and log-in credentials. Each letter will supply a telephone number to complete a CATI interview or get assistance with the self-administered web questionnaire. Parent letters will also request that they encourage their teenager to participate in the College Update, though both letters will mention that either the student or a parent can complete the questionnaire. Regardless of student's age, the student letters will provide the same information as the parent letters, albeit with different log-in credentials.

For the College Update, the field test data collection period will be divided into three phases and target incentives in a responsive design aimed at reducing bias in the final estimates. The three phases are:

- (1) A two-week web data collection period. At the start of the first phase of data collection, each of the parents and students in the College Update sample⁶ will receive a letter asking them to log onto the web to complete the questionnaire.
- (2) A three-week web plus CATI data collection period. After the two-week web-only data collection period, outbound calling to sample members will commence and continue for three weeks. These calls will be primarily directed to the student sample member. However, if the student is not available at the time of the call, callers will attempt to identify the parent who will be asked to complete the College Update questionnaire over the phone. If either the student or parent prefers to complete an online questionnaire, the interviewer will provide the parent with the study URL and log-in credentials. The interviewer will also offer to send the log-in information in an e-mail.
- (3) A nonresponse follow-up period. After three weeks of CATI data collection, the Mahalanobis distance function (discussed further below) will determine the target cases for nonresponse follow-up. Target cases will receive a \$5 pre-paid cash incentive in this reminder mailing, and the letter will promise that an additional \$10 will be sent upon completion of the questionnaire. Cases not identified as cases of interest for nonresponse follow-up will receive no monetary incentive.

Based on experience to date with HSLS:09 parent- and out-of-school student- data collection, an estimated 25 percent of sample members will participate by the end of the five-week early data collection period. Of the approximately 754 cases in the College Update Field Test, this leaves potentially 566 as estimated 'late' responders. Mahalanobis distance functions will be evaluated and a logical cut point (the largest distance scores) will be established so the incentive will be offered to approximately 375 of the 566 cases with the largest distance scores.

Transcripts. Transcript data will be requested for students who participated (or were questionnaire-incapable) in either of the in-school rounds of HSLS:09, from all schools attended since they first entered the sample in the 2009-10 (main study) or 2008-09 (field test) academic year. A complete transcript from the school will be requested as well as complete transcripts from transfer schools that the students attended, as applicable. The success of the transcript collection is closely tied to the active participation of selected schools. The consent and cooperation of a school's coordinator is essential and helps to encourage the timely completion of the transcript collection. If the HSLS:09 coordinators have been involved with the in-school collection, they will

⁶ The CU sample includes only those students who participated in the base year study and/or the first follow-up study. Students who did not participate in either prior round would not be included in the College Update or subsequent data collection rounds. Of the 754 students/parents included in the College Update field test sample, 26 sample members (3 percent) either requested that they be removed from the study or did not have sufficient contact information.

be familiar with HSLs:09 and recognize the study's importance. Procedures for working with schools will build upon the rapport developed with schools in the HSLs:09 base-year and first follow-up and will be based on successful past procedures. For example, HSLs:09 will use a transcript control system (TCS) similar to the system used for the Baccalaureate and Beyond Longitudinal Study 2008/09 (B&B:08/09) and Beginning Postsecondary Students 2004/09 (BPS:04/09) transcript collections to maintain relevant information about the schools attended by each cohort member and mail and telephone contacts made.

The descriptive materials sent to schools will be clear, concise, and informative about the purpose of the study and the nature of subsequent requests and will include letters from RTI and NCES and instructions for how to log on to the study's secure website and access information and tools for providing transcripts. Follow-up calls will be made to ensure receipt of the request packet and answer any questions about the study. It is likely that telephone prompting will be required to obtain the desired number of transcripts in addition to e-mail prompts, letters, and postcard prompts.

A seasoned team of Institutional Contactors (ICs) will be assigned a set of schools that is their responsibility throughout the transcript collection process. This allows Institutional Contactors to build and maintain rapport with school staff and to provide a reliable point of contact at RTI. Institutional Contactors will be thoroughly trained in transcript collection and in the purposes and requirements of the study, which helps them establish credibility with the school staff.

Different options for collecting transcripts for sampled students are offered. Data security procedures for each method of transcript collection will be addressed in the HSLs:09 Data Security Plan. The school coordinator is invited to select the method of greatest convenience to the school. School staff will have the option to provide transcript data by: 1) uploading electronic transcripts for sampled students to the secure study website; 2) sending electronic transcripts for sampled students by secure File Transfer Protocol; 3) sending electronic transcripts as encrypted attachments via email; 4) for schools that already use this method, RTI requesting/collecting electronic transcripts via a dedicated server at the University of Texas at Austin; 5) transmitting transcripts via a secure electronic fax at RTI, after sending a confirmed test page; and as a last resort 6) sending transcripts via an express delivery service after redacting personally identifying information. The majority of schools will likely fax the data, followed closely by those schools that will use FedEx. The numbers will be small for the other modes, but the plan is to set up multiple means to accept the data if the school is willing and able to use the more sophisticated electronic modes. More options

promote the likelihood of more timely response. For reference, the recent B&B/BPS Postsecondary Education Transcript Study (PETS) found 66% of the data arrived via fax from colleges. The percentage could be even higher with high school transcripts.

Consent Procedures. Privacy and consent concerns may arise in the collection of high school transcripts, particularly when transcripts are requested directly from high schools where third-party student transcript requests are likely uncommon. ICs requesting transcripts will be familiar with the Family Educational Rights and Privacy Act of 1974 (FERPA), which permits schools to release student data to the U.S. Department of Education and its authorized agents without consent, and will be prepared to respond to concerns raised by high school staff. If the school requires student consent to release the transcripts, RTI will prepare and mail consent forms to the students (or parents if the student is known to be under age 18). Consent forms should be returned directly to RTI, where the consent forms will be packaged and sent to the school with a second request for transcripts. Telephone prompting will be conducted as needed to remind students and parents to send consent forms to RTI. During the field test, the prevalence of schools requiring implied or explicit consent and the rate of return will be evaluated so that procedures may be refined for the main study. During the ELS:2002 high school transcript collection, 5 percent of the schools required explicit consent (i.e., signed consent form) to release transcripts.

In compliance with FERPA, a notation will be made in the student record that the transcript has been collected for use in HSLs:09.

B.4 Study Contacts

Laura LoGerfo and Jeff Owings are the primary contacts for the HSLs:09 study at NCES. Exhibit B-1 provides the names of contractor-affiliated consultants on statistical aspects of HSLs:09.

Exhibit B-1. Consultants on Statistical Aspects of HSLs:09

Name	Affiliation	Telephone
James Chromy	RTI	(919) 541-7019
Steven J. Ingels	RTI	(202) 974-7834
Jill A. Dever	RTI	(202) 974-7846
Daniel J. Pratt	RTI	(919) 541-6615
John Riccobono	RTI	(919) 541-7006
David Wilson	RTI	(919) 541-6990