

**Program for the International  
Assessment of Adult Competencies  
(PIAAC)  
2017 National Supplement**

**OMB# 1850-0870 v.5**

**Supporting Statement Part B**

**Submitted by:**

**National Center for Education Statistics  
U.S. Department of Education  
Institute of Education Sciences  
Washington, DC**

**July 2016**

# **COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS**

## **B.1 Importance of Information**

The PIAAC target population consists of non-institutionalized adults who at the time of the survey reside in the U.S. (whose usual place of residency is in the country) and who at the time of interview are between the ages of 16 and 74 years, inclusive. Adults are to be included regardless of citizenship, nationality, or language. The target population excludes persons not living in households or non-institutional group quarters, such as military personnel who live in barracks or bases, or persons who live in institutionalized group quarters, such as jails, prisons, hospitals, or nursing homes. The target population includes full-time and part-time members of the military who do not reside in military barracks or military bases, adults in other non-institutional collective dwelling units, such as workers' quarters or halfway homes, adults living at school in student group quarters, such as a dormitory, fraternity, or sorority. Persons who are temporarily in the U.S., for example on a work visa, will be included in PIAAC if they have lived in the sampled dwelling unit for 6 months or more of the last 12 months or if they have just moved into the dwelling and expect it to be their "usual residence" moving forward (where they will reside for 6 months or more per year). Adults who are unable to complete the assessment because of a hearing impairment, blindness/visual impairment, or physical disability are in-scope; however because the assessment does not offer accommodations for physical disabilities, they are excluded from response rate computations. Based on the most recent Census Population Estimates (2014), there are about 233 million people in the target population. The overall sample size goal for the PIAAC 2017 National Supplement is to reach 3,800 16-74 year olds, of which 3,500 are 16-65 year olds and 300 are 66-74 year olds. The reason for the two groups is that the international PIAAC comparison group is 16-65 year olds, and 66-74 is of special interest to the U.S.

Two prior U.S. PIAAC data collections have occurred to date. The United States completed its PIAAC 2012 Main Study along with the 23 countries who participated internationally. It included a sample of 5,010 adults in 80 Primary Sampling Units (PSUs). The survey components included a screener, an in-person background questionnaire, and a computer-based or paper assessment. A second U.S. data collection, the PIAAC 2014 National Supplement of 3,660 respondents, was conducted with the same survey components. The PIAAC 2012 and 2014 data collection efforts together formed an enhanced 2012/2014 combined nationally representative sample. The PIAAC 2017 National Supplement will be the third data collection for PIAAC in the U.S. National estimates of proficiency in literacy, numeracy and problem solving from the 2017 national sample will provide an additional set of data points for evaluating trends in proficiency as compared to estimates from the 2012/2014 combined sample, and earlier adult literacy surveys. In addition, as discussed further below, the data collected from all survey years (2012/2014/2017) will be used to form sub-national model-based estimates through small area estimation (SAE) methods.

## **B.2 Statistical Methodology**

This section describes the sample design for the PIAAC 2017 National Supplement. There are two core objectives for the sample design. First, the sample will be designed to ensure a nationally representative sample of the U.S. adult population 16 to 74

years old. Second, the sample will be designed to arrive at sufficient coverage of different types of counties in the United States, that, when combined with the PIAAC 2012 Main Study and the PIAAC 2014 National Supplement samples, can produce indirect small area estimates (e.g., counties) for the U.S. With an expected 3,800 respondents in the 2017 survey, the combined sample size for small area estimation will include about 12,470 individuals.

The sample design plan is based on the expectations that the SAE approach will be the same as the one used for the 2003 National Assessment of Adult Literacy (NAAL), as described in Mohadjer, et al. (2009, 2012), and as summarized in the “Estimation” section of this document. For PIAAC, we evaluated the feasibility of deriving county-level indirect estimates from a combined 2012/2014/2017 sample. The evaluation led to a sample size of 80 PSUs to field for the 2017 survey. In the evaluation, we first compared the county-level estimates obtained from models based on different number of sampled counties using the data from the 2003 NAAL. The goal was to evaluate whether the 92<sup>1</sup> counties in PIAAC 2012/2014 are sufficient to support the production of county-level estimates. The results were mixed, but tended to indicate a larger number of sampled counties are needed for this estimation. Next, we evaluated the coverage of the PIAAC 2012/2014 county sample as compared to the entire frame of counties in the U.S. In the analysis, the country was divided into 216 cells using a cross-classification of four county-level percentages of demographics related to proficiency. The cells covered by the 92 counties in the current PIAAC included 1,401 counties, which is about 45 percent of all counties in the U.S., while for comparison, the cells covered by the 264 counties in the larger 2003 NAAL SAE model included 2,165 counties, which is about 69 percent of counties. There are 47 cells not covered by PIAAC (mostly groups of small counties) that are covered by NAAL. This analysis concluded that the 2017 sample should bring in cases from a sample of about 60 new counties (from about 80 PSUs) in addition to the 92 counties already in the PIAAC sample. The new counties should maximize the number of counties from cells not currently covered by PIAAC at the same time as producing an optimum national sample for PIAAC 2017.

## **Sample Selection**

A four-stage, stratified area probability sample is planned that involves the selection of (1) primary sampling units (PSUs) consisting of counties or groups of contiguous counties, (2) secondary sampling units (referred to as segments) consisting of groups of blocks, (3) dwelling units (DUs), and (4) eligible persons (ultimate sampling unit) within DUs. Random selection methods will be used, with calculable probabilities of selection at each stage of sampling. The sample selection approach is described for each sampling stage below in turn.

For the initial stage of sampling, a total of 80 PSUs will be selected. The PIAAC 2017 National Supplement PSUs will be formed in the same way as the PIAAC 2012 Main Study and the PIAAC 2014 National Supplement<sup>2</sup>. A stratified probability-proportionate-to-size (PPS) sample will be selected, where the measure of size (MOS) is the estimated

<sup>1</sup> PIAAC 2012/2014 selected 80 Primary Sampling Units, which resulted in respondents within 99 counties. However, direct estimates were possible for only 92 counties due to small sample size in 7 of the counties.

<sup>2</sup> In PIAAC 2012 and 2014 the PSUs were formed by combining adjacent counties to reach a minimum population size, respecting state and metropolitan statistical area boundaries, and taking into consideration the travel distance for data collectors.

non-institutionalized population—adjusted from the resident population estimates from the 2014 Census Bureau population estimates<sup>3</sup> available for each county. The PSUs with the largest MOS will be selected with certainty (with probability equal to one) before stratification using a certainty cutoff determined from PPS sampling. One PSU will be selected per stratum, where strata will be formed from county-level variables relating to metropolitan statistical area status, race/ethnicity, poverty, foreign born status, and education attainment<sup>4</sup>. Strata will be close-to-equal in size in order to reduce the variation in workload and also to control the variances of the estimates. County data is available from the Census Bureau’s Population Estimates Program and the American Community Survey.

To support the second core objective of producing indirect county-level estimates, the PSUs were selected in such a way as to reduce the number of PSUs that are in both the PIAAC 2012 Main Study/PIAAC 2014 National Supplement and PIAAC 2017 National Supplement samples. Minimizing the overlap will help maximize the coverage of the combined sample across demographic variables and optimize county-level estimation given the available sample size. The sampling procedure called Ohlsson’s method (Ohlsson (1996)) was used to achieve this goal. Ohlsson’s algorithm uses permanent random numbers to reduce overlap between two or more samples. Ohlsson’s method is applicable to probability proportionate to size (PPS) samples with one unit selected per stratum. The result of its’ application had a large impact on the coverage of the cells and new counties selected<sup>5</sup>, while arriving at an optimum design for a nationally representative sample of US adults for 2017, at the same time as arriving at sufficient coverage of different types of counties for small area estimation using a combined 2012/2014/2017 sample.

Four PSUs with the largest measures of size (MOS), as defined by the population 15 to 74 years of age (slightly different age range available from the 2014 Census Population Estimates) were selected with probability equal to one before stratification using a certainty cutoff determined from probability proportionate to size sampling. Such PSUs are referred to as self-representing. The non-self-representing PSUs on the frame were grouped into major strata. The major strata were based on whether the PSU was part of a metropolitan area. Once major strata were identified, substrata (minor strata) were formed via a nested stratification process, as discussed in Krenzke and Haug (2009), using auxiliary variables related to the expected proficiency scores. The county-level variables used for stratification were related to the county-level variables used for the NAAL SAE task (Mohadjer et al. 2009): race/ethnicity, poverty, foreign-born status, and education attainment. Strata were close to equal in size to reduce the variation in interviewer workload.

Once the strata were formed, one non-self-representing PSU was selected per stratum with probability proportionate to its MOS, while minimizing the overlap with the PSUs selected for the 2012/2014 PIAAC. The resulting 80 self-representing and non-self-

---

<sup>3</sup>U.S. Census Bureau Population Estimates Program produces estimates of the resident population at the county level. An adjustment will be done to estimate the non-institutionalized population for each county.

<sup>4</sup> These variables were found to be related to literacy proficiency based on a research conducted by Westat for a Small Area Estimation task using NAAL data (Mohadjer, et al., 2009).

<sup>5</sup> 90 new counties (not selected in 2012/2014) were selected in the PIAAC 2017 sample and 30 new cells were covered by the 90 counties.

representing PSUs are diverse in terms of literacy skills, geographic region of the country, and urbanicity of the PSU, as well as diverse in education attainment, foreign-born status, race/ethnicity, and poverty status.

For the second stage of sampling, we will select a PPS systematic sample of 684 segments from the 80 sample PSUs. The systematic selection will use a sorted list based on the geographic sequencing of the SSUs within the PSU in order to ensure spatial representation, which also improves the coverage of demographic subgroups. The average SSU size is planned to be about 150 to 200 dwelling units, while the average SSU size (used for the PIAAC 2012 and 2014 samples) was about 100 dwelling units. The larger SSU sizes will have a decreasing impact on the variances, and therefore a smaller number of SSUs is needed in each PSU. The use of purchased postal addresses (as described in the third stage of sampling below), as opposed to traditional listing operations, allows for larger secondary sampling units to be used.

The third stage of sampling for PIAAC will involve an initial sample of about 6,626 dwelling units (DUs) from the frame of addresses in each selected segment in order to arrive at about 3,800 completed assessments. To form the DU sampling frames within sampled SSUs, we plan to employ address-based sampling (ABS) for the PSUs that have high-quality ABS lists and use the traditional listing procedures in other PSUs. ABS utilizes residential addresses from the most recent U.S. postal service (USPS) computerized delivery sequence file (CDSF). The USPS address lists have been shown to have around a 98% coverage rate overall; however, this can be considerably lower for some rural areas. Therefore, we will use ABS for the PSUs that have high-quality ABS lists and use the traditional listing procedures in other PSUs. Traditional listing will be carried out by trained Westat listers. Once the DU sampling frames are constructed, DUs will be selected systematically in each SSU, where the DUs will be sorted geographically (by ZIP code and carrier route information for the USPS lists, and by listing order for the list compiled from traditional listing) to best avoid sampling neighboring households.

To reduce the potential for coverage issues for PSUs that employ ABS, we will implement an address coverage enhancement (ACE) approach that permits representation of all addresses. We have developed the ACE methodology when using USPS list frames for household sampling that tackles both the coverage and geocoding issues while taking full advantage of the electronic nature of the CDSF, and permits representation of all addresses while limiting the resource requirements (Dohrmann et al. (2012), Kalton, Kali and Sigman (2014)). The ACE is similar to the procedure conducted in the PIAAC 2012 Main Study and the PIAAC 2014 National Supplement samples to identify any DUs not captured in the original frame created through traditional listing.

The fourth stage of selection involves a two-phase sampling approach, which is implemented through a screener questionnaire, first, to collect information about persons in the household, and second, to determine eligibility and stratify by age group, and to apply different sampling rates to arrive at the target sample sizes in each age group. The screener begins by listing the age-eligible household members (aged 16 to 74) in the selected dwelling unit during the screener interview. Once persons are enumerated in a household, two strata will be formed, 1) 16-65 year olds, and 2) 66-74 year olds. In stratum 1), one person will be selected if there are three or less in the household, otherwise two persons will be selected. Selecting two persons in the larger households helps to reduce the variation due to unequal probabilities of selection,

which has an increasing effect on the resulting variance estimates. For stratum 2), three out of every five households with at least one 66-74 year old will be flagged to select one 66-74 year old within the household. This subsampling is needed to keep the sample size of 66-74 year olds at 300, but will cause a very slight increase in the design effect. Households without an individual in either of the two strata will 'screen out' (that is, classified as 'ineligible').

The enumeration and selection of persons will be performed using the CAPI system, which will collect information via the screener instrument, including age and gender of persons in the dwelling unit, and randomly select eligible respondents.

Household members who are staying at college dormitories will be considered to be part of their family's household. If it is not possible to reach the students at the family homes during the data collection period, an interview will be arranged with them at college, if they reside within or adjacent to one of the 80 PSUs. Westat successfully applied the same procedure for the PIAAC 2012 Main Study and the PIAAC 2014 National Supplement.

Table 5 below provides a summary of the sample sizes needed at each sampling stage to account for sample attrition due to (a) ineligibility (i.e., households without at least one 16-74 year old adult or vacant dwelling units), (b) screener nonresponse, (c) within-household selection rates, and (d) nonresponse to the BQ and the assessment. The occupancy rate is expected to be about 85 percent, similar to what was experienced in the PIAAC 2012 Main Study. The response rates are consistent with the corresponding weighted response rates for the PIAAC 2012 Main Study sample except for the assumed BQ and assessment response rates for the 66-74 year olds, which are based on the 66-74 year olds in the 2014 household sample (the 2012 sample did not include 66-74 year olds). Overall, a 70 percent response rate is expected for the 16-65 year olds and 60 percent response rate for the 66-74 year olds. If the actual response rates do not meet the National Center for Education Statistics (NCES) standards for response rate goals, a nonresponse bias analysis will be conducted at each stage of data collection that do not meet the standards (see Appendix F for preliminary plans for conducting these analyses). The ineligibility rate (due to households without at least one person in the target population) is based on the public use microdata sample (PUMS) from the 2014 American Community Survey. The expected percentage of households with two sample persons is based on both the 2014 ACS PUMS and selection rules established for PIAAC.

In addition, this sample will be increased to provide a reserve sample of households. The reserve sample will only be used in case there are unusual response problems or unforeseen sample losses observed during the data collection. A reserve sample of about 50 percent the size of the main sample will be selected randomly and set aside to be used in case of a shortfall in the sample.

## **Estimation**

After data collection, sampling weights will be produced to facilitate the estimation of the target population parameters for the PIAAC 2017 national estimates. Replicate weights will be computed to facilitate variance estimation, and will capture the variation due to the sample design and selection, as well as weighting adjustments. We will also analyze the nonrespondents and provide information about whether and how they differ from the respondents, as required by NCES statistical standards.

For the purposes of small area estimation, the 2012/2014 combined sample will be brought together with the 2017 sample to improve the precision of the resulting small area estimates. Composite weights were produced in the 2012/2014 combined sample so that national estimates could be generated for the combined sample. Similarly, composite weights ( $\widetilde{W}_i^\square$ ) will be created from the two sets of weights (from the 2012/2014 combined sample, and the 2017 sample) as follows:

$$\widetilde{W}_i^\square = \alpha W_i^{2012/2014} I_{2012/2014}(i) + (1 - \alpha) W_i^{2017} I_{2017}(i),$$

where the term  $\alpha$  is the compositing factor,  $W_i^{2012/2014}$  is the 2012/2014 final weight for person  $i$ ,  $W_i^{2017}$  is the 2017 final weight for person  $i$ . The indicator variable  $I_{2012/2014}(i) = 1$  if  $i$  is in the 2012/2014 combined sample,  $= 0$ , otherwise; and the indicator variable  $I_{2017}(i) = 1$  if  $i$  is in the 2017 sample, or 0 otherwise.

This method produces unbiased estimates for any value of the compositing factor. The optimum value is the one that results in the lowest variance. For a particular estimate  $Y$ , the optimum value would be calculated as:

$$\alpha = \frac{V(\widehat{Y}_\square^{2017})}{V(\widehat{Y}_\square^{2012/2014}) + V(\widehat{Y}_\square^{2017})},$$

where  $V(\widehat{Y}_\square^{2012/2014})$  represents the variance of the estimate of  $Y$  in the 2012/2014 combined sample and  $V(\widehat{Y}_\square^{2017})$  represents the variance of the estimate of  $Y$  in the 2017 sample. Further discussion of combining samples is provided in the 2003 NAAL Technical Report in the context of combining several independent state samples with the national sample. Other discussions as it relates to dual-frame estimation can be found in Lohr (2011).

Small area estimates will be generated for all counties and states in the U.S. NCES has produced state and county level estimates of the percentages of adults lacking Basic prose literacy skills, and Westat expects to follow those same procedures for PIAAC. The model used was a Hierarchical Bayes (HB) area-level model based on the 2003 NAAL and auxiliary data from the 2000 Census (see for example Mohadjer et al. 2009). A multi-stage sample design was employed for the NAAL with counties or groups of counties selected as the PSUs. Only counties within sampled PSUs had any sample cases in the NAAL.

The aim of the NAAL SAE model was to estimate the true percentage of adults who are lacking Basic prose literacy skills (as evaluated by the NAAL instrument) within county  $i$  in state  $s$ , given as  $p_{si}$  below. The unbiased estimator of  $p_{si}$  within a county with NAAL data is  $\widehat{p}_{si}$ , which is subject to both sampling and measurement error<sup>6</sup>, combined into a single error term  $e_{si}$ . The dependent variable  $z_{si}$  in the mixed-effects regression model is the logit of the true percentage  $p_{si}$  (the logarithm of the odds, where the odds is the ratio  $\frac{p_{si}}{1 - p_{si}}$ ). The following sampling model and mixed effects regression linking model were used in NAAL:

$$\widehat{p}_{si} = p_{si} + e_{si}$$

$$z_{si} = \beta_0 + \beta_1 x_{1si} + \beta_2 x_{2si} + \beta_3 x_{3si} + \beta_4 x_{4si} + \beta_5 x_{5si} + \beta_6 x_{6si} + v_s + u_{si}$$

<sup>6</sup> The measurement error is from the prose literacy test.

where,  $z_{si} = \ln\left(\frac{p_{si}}{1-p_{si}}\right)$ , and

where  $\beta_0$  through  $\beta_6$  are fixed regression parameters,  $v_s$  is a state random intercept representing the difference between the true value of the characteristic for the state and its model-based expectation and  $u_{si}$  is a corresponding county random intercept, and the six predictor variables are denoted by  $x_{jsi}$ , where  $j=1, \dots, 6$ , and are described as follows:

- Percentage of the population who are foreign-born (and in the U.S. less than 20 years);
- Percentage of age 25 or older adults with a high school education or less;
- Percentage of Blacks and/or Hispanics;
- Percentage of persons below 150 percent of the poverty line;
- A dichotomous zero to one indicator for two Census divisions (New England and North Central).
- A dichotomous indicator for a State Assessment of Adult Literacy state<sup>7</sup>.

The random effects  $v_s$ ,  $u_{si}$ , and  $e_{si}$  were assumed to be normally distributed with mean zero and variances  $\sigma_v^2$ ,  $\sigma_u^2$ , and  $\sigma_e^2$  respectively, and were assumed to be independent. The variance  $\sigma_e^2$  was estimated from the NAAL.

The six predictor variables listed above were chosen after an assessment of extensive analyses of a large number of possible predictors. The larger set included county information from the 2000 Census and state information from the ACS and Census projections. This larger set was chosen as variables that were thought to be potentially correlated to literacy levels. The process of selecting the final variables from this larger set was based on correlation analyses and stepwise logistic regression procedures. Mohadjer et al. (2009) provide a complete listing of the initial pool of variables and provide details of the variable selection process.

The NAAL small area program used an HB approach (see for example Rao 2003, Chapter 10) rather than the simpler Empirical Bayes approach, because of the use of the logit link in the regression model. This model is an 'unmatched model' as opposed to the Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program case of a 'matched model'<sup>8</sup>. In the unmatched model case, as introduced by You and Rao (2002), there is an intermediate function between the unbiased sample estimate and the linear predictive model. The intermediate function (logit) was necessary due to small sample size as well as the low estimated proportions. As with the Empirical Bayes approach, the final indirect estimates at the county level are combinations (not necessarily linear) of model predictions and direct estimates for counties with NAAL survey data, and are entirely model predictions for counties without NAAL survey data. The state-level estimates are weighted averages of the county-level estimates.

The estimation procedures for the PIAAC data are prescribed by and are the responsibility of the international sponsoring agency; however, the United States has reviewed and agrees with these procedures. The United States will comply with these

---

<sup>7</sup> These were states which had separate state literacy studies (partially funded by the state).

<sup>8</sup> More information available at (<http://www.census.gov/hhes/www/saipe/>)



procedures and policies by delivering masked data (note that a disclosure analysis will be conducted prior to submitting the data to the international contractor so as to comply with current federal law), and documentation of sampling and weighting variables. All data delivered to the PIAAC Consortium will be devoid of any data that could lead to the identification of individuals.

### **Degree of Accuracy**

The design is very similar to the PIAAC 2012 Main Study. Therefore, as was the case in the Main Study, a design effect of about 2 is expected for the PIAAC 2017 National Supplement. Based on the PIAAC 2012 Main Study results, Standard errors for national estimates of average literacy scores are expected to be about 1.0 for scores of magnitude about 270. By education achievement, for example, the standard error for average literacy scores are expected to be about 2 for groups relating to low, medium, and high education achievement.

### **Specialized Sampling Procedures**

To reduce costs, in lieu of traditional listing, the sampling frame of dwelling units in a majority of SSUs will be formed from purchased addresses based on files that originate from the U.S. Postal Service.

### **Any Use of Periodic (Less Frequent Than Annual) Data Collection Cycles to Reduce Burden**

There are no anticipated problems that would require specialized sampling procedures, nor will there be any use of periodic data collection cycles to reduce burden.

**Table 5. PIAAC 2017 National Supplement sample yield estimates**

<b>Survey and sampling stages</b>	<b>Eligibility and response rates</b>	<b>Projected rates</b>	<b>Expected sample size</b>
Number of selected PSUs			80
Number of selected SSUs			684
Expected number of selected households (HHs)	Occupied dwelling unit rate	85.0%	6,626
Expected number of occupied dwelling units	Screening response rate	86.5%	5,632
Expected number of completed screeners			4,872
Expected number of eligible screeners	Eligibility rate 16-65 (66-74)	83.2% (9.0%)	4,055 (438)
Expected number of attempted BQs	Percentage of HHs with two sample persons 16-65 (66-74)	6.6% (0%)	4,325 (438)
Expected number of persons with completed BQs	BQ response rate 16-65 (66-74)	82.5% (70.5%)	3,568 (309)
Expected number of completed or partially completed assessments	Assessment completion rate 16-65 (66-74)	98.1% (97.7%)	3,500 (300)

NOTE: Figures in parentheses are for the subgroup 66-74 year olds.

## **B.3 Maximizing Response Rates**

NCES views gaining respondent cooperation as an integral part of a successful data collection effort and will invest the resources necessary to ensure that the procedures are well developed and implemented. We will use an advance contact strategy that has been successfully employed on many large-scale, in-person household studies. An advance letter will be mailed to all households selected for the household-based sample in advance of the data collector's initial visit. This letter will inform potential respondents of NCES authorizing legislation; the purposes for which the PIAAC data are needed; uses that may be made of the data; and the methods of reporting the data to protect privacy. In addition, an informative brochure (provided in Appendix E) will be given to sampled participants when the interviewer visits the sampled household. All project materials will include the study's web site address and a toll-free telephone number for respondents to obtain additional information about the study. The materials will also mention the respondent incentive and will include the study logo for legitimacy purposes. It is very important for the data collector to establish legitimacy at the door, which can be accomplished by the use of a strong introductory statement during which the data collector shows their ID badge and a copy of the advance materials.

Once data collection begins, effective contact patterns are another important component of achieving response rates. Completion rates improve when data collectors attempt contact on different days of the week and at varying times of the day. Data collectors will make four well-timed attempts to contact a household before reviewing the case with the supervisor to identify another pattern of contact. These other contact strategies may include telephone, FedEx letters, or leaving messages with neighbors. We plan to staff each PSU with two data collectors. It is advantageous to have multiple data collectors in a PSU as it allows better matching between data collectors and respondents and allows for coverage in case of data collector illness or unavailability. In carrying out efforts to achieve high response and participation rates, we propose to organize our data collection efforts using a phased approach that allows for refusal conversion.

Each data collector will receive a laptop computer loaded with the Interviewer Management System. This system allows data collectors to launch all CAPI instruments and permits tracking of their work and time. Data collectors will use the electronic record of call (EROC) feature of the Interviewer Management System to collect information about each visit to a household that did not result in a completed interview. EROC information will include: contact date and time, contact result or disposition code, appointment information, and general data collector comments. The EROC data are very helpful in documenting the results of contact attempts for nonresponding households, and in helping to design a more directed and effective campaign to convert the nonresponding households. All nonresponse followup and refusal conversion efforts also will be tracked and documented in the Interviewer Management System.

Whenever a refusal or breakoff is encountered, the data collector will complete an automated noninterview report (NIR) that captures information about the reason for refusal. Automated EROC and NIR information is available to the supervisors via data transmission to the home office by the data collectors and subsequent transmissions to the supervisors. Contact and decline information will be collected, coded, and included in the biweekly data collection progress report. NCES believes that frequent, open

communication between all levels of field staff is required for a successful data collection effort. Supervisors will primarily use email for day-to-day communication with their staff. Scheduled weekly conference calls will also be used at all levels. All supervisory staff will be available for questions or other issues that come up every day via telephone and email. Other activities that will be considered to increase response rates are:

- **Enhance interviewer training on screening.** Dedicate more training time, in the home study package and during initial interviewer training, to “the importance of obtaining high screener completion rates and tips on completing screeners.” Continue to focus on this throughout data collection via supervisor/interviewer conference calls. Finally, hold special conference call training sessions, as necessary during data collection, to focus on this activity.
- **Design an interviewer incentive program.** For most of the PIAAC 2012 Main Study and the entire PIAAC 2014 National Supplement, Westat had an interviewer incentive program that rewarded completes. For the PIAAC 2017 National Supplement, we would plan to have an incentive program for interviewing/assessment work from the first day of field work.

## **B.4 Purpose of Field Test and Data Uses**

The same procedures, instruments, and assessments that were used for the PIAAC 2012 Main Study and the PIAAC 2014 National Supplement will also be used to conduct the PIAAC 2017 National Supplement.

## **B.5 Individuals Consulted on Study Design**

The following are responsible for the statistical design of PIAAC:

- Holly Xie, National Center for Education Statistics;
- Stephen Provasnik, National Center for Education Statistics;
- Leyla Mohadjer, PIAAC Consortium/Westat; and
- Kentaro Yamamoto, PIAAC Consortium/Educational Testing Service.

The following are responsible for sampling activities:

- Leyla Mohadjer, PIAAC Consortium/Westat; and
- Tom Krenzke, PIAAC Consortium/Westat.

Analysis and reporting will be performed by:

- Kentaro Yamamoto, Educational Testing Service.

## **References**

Krenzke, T., and Haung, W. (2009). Revisiting Nested Stratification of Primary Sampling Units. *Proceedings of the Federal Committee on Statistical Methodology*. Retrieved from [www.fcsm.gov/events/index.html](http://www.fcsm.gov/events/index.html).

Lohr, S.L. (2011). Alternative Survey Sample Designs: Sampling with Multiple Overlapping Frames. *Survey Methodology* 37(2): 197–213.

Mohadjer, L., Kalton, G., Krenzke, T., Liu, B., Van de Kerckhove, W., Li, L., Sherman, D., Dillman, J., Rao, J.N.K., and White, S. (2009). *Indirect county and state estimates of the percentage of adults at the lowest level of literacy for 1992 and 2003* (NCES 2009-482). Washington, DC: Institute of Education Sciences, U.S. Department of Education.

- Mohadjer, L., Rao, J.N.K., Liu, B., Krenzke, T., and Van De Kerckhove, W. (2012). Hierarchical Bayes small area estimates of adult literacy using unmatched sampling and linking models. *Journal of the Indian Society of Agricultural Statistics*, 66(1), 55-64.
- Ohlsson, E. (1996). *Methods for PPS size one sample coordination*. Stockholm, Sweden: Institute for Actuarial Mathematics and Mathematical Statistics, Stockholm University, No. 194.
- Rao, J.N.K. (2003). *Small area estimation* (Wiley Series in Survey Methodology). New York: Wiley.
- You, Y., and Rao, J.N.K. (2002). Small area estimation using unmatched sampling and linking models. *Canadian Journal of Statistics*, 30(1), 3-13.