

**PROGRAM FOR THE INTERNATIONAL
ASSESSMENT OF ADULT COMPETENCIES
(PIAAC)
2011-2012
MAIN STUDY DATA COLLECTION**

**REQUEST FOR OMB CLEARANCE
OMB# 1850-0870 v.2**

Supporting Statement Part B

Prepared by:

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

March 1, 2011

Collection of Information Employing Statistical Information

A large blue square containing a white letter 'B' in a bold, sans-serif font.

B.1 Respondent Universe and Response Rates

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 65 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.

The main study assessment will be comprised of a probability-based nationally representative sample of 5,000 persons. The standard PIAAC design requires random selection methods with calculable probabilities of selection at each stage of sampling for the main study. Thus each person in the target population will have a known non-zero probability of selection. A four-stage sample design will be employed in which the primary sampling units (PSUs) will be counties or groups of contiguous counties. The second stage will be segments (census blocks or combinations of blocks), the third stage will be dwelling units (DUs), and the fourth stage will involve selecting one or two eligible adults per household. Once dwelling units are selected, a screener interview will be conducted to identify the eligible persons within selected households. A sampling algorithm will be implemented within the CAPI system to select one or two sample persons among those identified to be eligible. Once selected, the background questionnaire (BQ) interview is to be completed. Upon completion of the BQ, the selected person will answer the Core Task items. If the respondent passes the Core Task, the respondent

will be provided a computer-based assessment. Those who do not pass the Core Task will be given a paper-and-pencil assessment booklet.

Main Study Sample Sizes and Response Rates

For the main study, an initial sample size of about 9,600 dwelling units is derived to ensure that the target number of completed assessments (5,000) can be achieved. The initial sample size needs to account for ineligibility (dwelling units without a person 16 to 65 years old and vacant dwelling units) and screener nonresponse, as well as nonresponse to the BQ module and the assessment. If the actual response rates do not meet the 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. (Note that the nonresponse bias analysis plan will be prepared in June 2011, per the international schedule, and will be submitted to OMB at that time.) Given the results of the incentive experiment and the experience in the PIAAC field test, we expect the overall response rate to be 68 percent if the incentive level is \$35, which is at the same level as the 2003 Adult Literacy and Lifeskills (ALL) survey, and to be 74 percent if the incentive level is \$50.

The occupancy rate is expected to be about 85 percent. This 15 percent vacancy rate is slightly higher than the actual 2003 ALL vacancy rate (13.6 percent) and the Census Bureau's American Community Survey (about 12 percent for 2005-2007), and at the same level as observed in the 2003 National Assessment of Adult Literacy and in recent years for the National Health and Nutrition Examination Survey conducted at Westat. The screener eligibility rate of 85 percent is the estimated proportion of households that have at least one individual between 16 and 65 years old, inclusive, from the October 2010 Current Population Survey. Table 5 provides a summary of the sample sizes and the response rate assumptions at each sampling stage.

Table 5. PIAAC Main Study: Sample yield estimates for 80 PSUs and 5,000 completed cases

Survey and sampling stages	Eligibility and response rates ¹	Project ed rates	Sample yield
Number of selected PSUs			80
Number of selected segments			900
Number of selected dwelling units			9,610
	Occupied dwelling unit rate	85.0%	
	Screening response rate	86.7%	
	Eligibility rate	84.9%	
	Percentage of Dwelling Units with Two Sampled Persons	6.0%	
Number of attempted BQs	BQ response rate	80.0%	6,371
Number of persons with completed BQs	Assessment completion rate	98.1%	5,097
Number of completed or partially completed assessments			5,000

¹ The screener, BQ and assessment response rates and occupied dwelling unit rate were determined based on experiences that include the PIAAC field test and the 2003 ALL. The eligibility rate and average number of sample persons per dwelling unit were computed from the 2010 Current Population Survey (a joint effort between the Bureau of Labor Statistics and the Census Bureau).

B.2 Procedures for Collection of Information

Statistical Methodology

This section describes the sample design for the PIAAC main study. A multi-stage design will be employed for the main study, and the sample selection approach is described for each sampling stage.

As mentioned in section B.1, the PIAAC target population consists of non-institutionalized adults 16 to 65 years old who reside in the United States at the time of interview. To arrive at a minimum of 5,000 completed cases, 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 area 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.

For the initial stage of sampling, a total of 80 PSUs will be selected. The PSUs will be 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. A stratified probability-proportionate-to-size (PPS) sample will be selected, where the measure of size (MOS) is the estimated non-institutionalized population—adjusted from the residency population estimates from the 2008 Census Bureau population estimates¹ 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 variables relating to census region, core-based statistical area status, race/ethnicity, poverty, English speaking ability, and education attainment. Westat conducted an extensive search for county variables for a Small Area Estimation (SAE) task using NAAL data (Mohadjer, et al., 2009), and the key predictors of literacy proficiency were related to race/ethnicity, poverty, English speaking ability, education attainment, and census division. 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 from other sources, including the American Community Survey and the Census Bureau’s Small Area Income and Poverty Estimates program.

For the second stage of sampling, we propose to select a PPS sample of 900 segments from within the 80 sample PSUs. The segments will consist of at least 60 dwelling units (DUs) in area blocks² (as defined by the 2000 census)

¹ 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.

² Blocks are very fine partitions of the United States, formed using visible semi-permanent features such as roads, railroad tracks, mountain ridges, bodies of water, and power lines. The only invisible boundaries used are county, state, and national boundaries. Minor civil division boundaries and property lines are ignored. A block group is a small group of contiguous blocks. A tract is a collection of contiguous block groups all within the same county.

or combinations of two or more nearby blocks. The frame of segments will be created within the selected PSUs using the Census 2000 Summary File 1 (SF1) block data.

Within each PSU, the block data from the SF1 files will be sorted by tract, block group, and block number, before creating the segments. Blocks with no DUs and no population will be included so that all areas, some of which may contain DUs constructed after the 2000 Census, will be involved in the formation process. Once segments are formed, the number of DUs in each segment will be compared with counts of residential addresses from the November 2010 Computerized Delivery Sequence File from the United States Postal Service (USPS). Segments for which the USPS counts indicate growth in the number of residential DUs will be investigated, and the number of DUs in the segment will be adjusted upward so that the sampling may be as accurate as possible.

The third stage of sampling for PIAAC will involve an initial sample of about 9,950 DUs from the frame of addresses in each selected segment in order to arrive at 5,000 completed assessments. All DUs within each selected segment will be listed by trained Westat listers. Given the actual number of listed DUs and derived sampling rates for each segment, dwelling units will be selected from the listing sheets at the home office. The listers will contact Westat whenever the number of listed DUs falls outside the range and provide any apparent reasons for the discrepancy. From the listings, the address and ID number of each selected DU will be keyed and verified.

The fourth stage of selection involves listing the age-eligible household members (aged 16 to 65) for each selected dwelling unit during the screener interview. Subsequently, one person will be selected at random within dwelling units with three or fewer eligible persons, and two persons will be selected if the dwelling unit has four or more eligible persons. 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. The design involves the selection of two persons in dwelling

units with a large number of eligible persons to prevent a substantial increase to the variation in the resulting sampling weights.

Household members who are away in college (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 field test PSUs. Westat successfully applied the same procedure for the 2003 ALL survey.

Estimation

For the main study, sampling weights will be produced to facilitate the estimation of the target population parameters. 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.

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 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.

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.

B.3 Maximizing Response Rates

In order to meet the PIAAC response rate goals, NCES will rely on procedures and approaches that have been used successfully over many years of conducting household studies. Building good response rates begins with hiring field staff with the experience and skills that will make them successful in convincing people to cooperate, and training them how to not only administer the instrument and follow the study procedures, but also how to convince respondents to participate.

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 in-person household studies. Advance materials, including a letter and an informative brochure (provided in Appendix D), will be mailed to all selected households in advance of the data collector's initial visit. These advance materials will inform potential respondents of NCES enabling 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 ensure confidentiality. 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.

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. We propose that data collectors 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 (IMS). 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 IMS 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 IMS.

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.

B.4 Tests of Procedures

The U.S. participated in a full field test for PIAAC. The PIAAC field test provided an opportunity for testing several facets of sampling. The main objectives of the sampling activities were to:

1. Provide a sample of adults that will be used to validate the test items to be included in the psychometric assessment;
2. Test the within-household sample selection process;
3. Train field staff in the sampling activities;
4. Test the Quality Control (QC) sampling-related procedures; and
5. Test the flow of materials and the sample data from sample selection to the delivery of the Sample Design International File (SDIF) at the end of the data collection.

B.5 Individuals Consulted on Statistical Design

The following are responsible for the statistical design of PIAAC:

- Leyla Mohadjer, PIAAC Consortium/Westat; and
- Kentaro Yamamoto, PIAAC Consortium/Educational Testing Service.

Westat will be the contractor responsible for sampling activities:

- Leyla Mohadjer, Vice President; and
- Tom Krenzke, Senior Statistician.

Analysis and reporting will be performed by:

- Kentaro Yamamoto, Educational Testing Service.