Programme for the International Assessment of Adult Competencies: U.S. Incentive Experiment

Tom Krenzke[1], Leyla Mohadjer[1], Hongsheng Hao[1]
[1] Westat, 1600 Research Boulevard, Rockville, Maryland 20850

Abstract:
Sponsored by the National Center for Education Statistics, the U.S. Programme for the International Assessment of Adult Competencies (PIAAC) field test data collection occurred between September and November 2010. Each participant was administered an in-person background questionnaire, and a literacy assessment. An experiment was designed to evaluate the impact of increasing the incentive amount from $35 to $50 to account for the added burden of a longer interview and assessment than past literacy surveys and the increased complexity of the PIAAC computer-based assessment. Incentive payments were randomly assigned to each segment (clusters of dwelling units within PSUs), so that each interviewer was assigned both incentive amounts to minimize any interviewer impact. The statistical analyses were conducted to examine differences between refusal rates, number of attempts and item response rates at the two incentive levels for the screener, background questionnaire (BQ), and screener and BQ combined. The analysis results provided support for an increase to the incentive amount for the PIAAC main study.

Key Words: Assessments, attempts, refusal rates, response rates

1. Introduction

The Programme for the International Assessment of Adult Competencies (PIAAC) is the most comprehensive international survey of adult skills ever undertaken. The survey is examining literacy in the information age and assessing adult skills consistently across the 23 participating countries. It is focusing on what are deemed key skills for individuals to participate successfully in the economy and society of the 21st century. As part of the efforts to meet PIAAC main study response rate goals, as sponsor of the survey, the National Center for Education Statistics (NCES) gave main study respondents a payment to thank participants for their time answering screener questionnaire items, background questionnaire items, and taking the assessment. NCES proposed to provide such a payment as an incentive to participants because (a) in recent years in-person household-based surveys have seen response rates decline, (b) research indicates that incentives play an important role in gaining respondent cooperation in such household surveys, and (c) PIAAC places a greater response burden on respondents than two similar surveys did, namely, the 2003 Adult Literacy and Lifeskills (ALL) or 2003 National Assessment of Adult Literacy (NAAL). Hence, PIAAC is at greater risk of respondents breaking off the questionnaire or assessment before they are both completed. This paper describes how the incentive amount was determined by an experiment conducted during the field test.

Many in-person household-based surveys have experienced decreasing response rates in recent years. For example, the National Health Interview Survey (NHIS), a one-hour interview, experienced a response rate decline of 12 percent from 1997 to 2007. The response rate for the National Survey on Drug Use and Health (NSDUH), decreased 5 percent between 2002 and 2007, and Round 1 of the Medical Expenditure Panel Survey (MEPS), which consists of a two-hour interview, sustained a response rate decline of 5 percent from 2001 to 2007.

In addition, the National Household Education Surveys (NHES) Program, which has collected information on important educational issues through telephone surveys of households in the United States since 1991, had response rates greater than 80 percent in 1991 and 1993, but in 1995 and 1996, they fell to 73 and 70 percent, respectively; in 2001 and 2003, they declined to 68 and 62 percent, respectively; and in 2007, they declined to 53 percent.

Research indicates that incentives play an important role in gaining respondent cooperation, especially in surveys that ask respondents to give several hours of their time and undertake a complex and often difficult assessment. A meta-analysis of 39 studies experimenting with incentives in telephone and in-person surveys from 1970 to 1997 (Singer, Van Hoewyck, Gebler, Raghunuthan, and McGonagle, 1999) found that incentives have a significant positive effect on response rates for both types of surveys. More specifically, they found that each dollar of an incentive paid resulted in approximately one third of a percentage point difference in response rate between the no incentive and the incentive conditions. Similar results were found for studies that had a low-incentive condition and a high-incentive condition. The effects found by the authors were linear, and therefore they concluded that "within the limits of incentives and response rates occurring in these experiments, more money results in higher response rates."

More specifically to literacy studies, a study was conducted to ascertain the effect of monetary incentives on response rates, among other variables (Mohadjer, Berlin, Rieger, Waksberg, Rock, Yamamoto, Kirsch, and Kolstad, 1997). The study included experiments with incentives in the National Adult Literacy Survey Field Test and Main Study. In both experiments, incentives produced a significant increase in response rate, most effectively in groups with low educational attainment and minority populations who are usually underrepresented in such studies. This effect would improve the distribution of these groups in the sample and therefore provide a better representation of the study's target population.

Recently, in 2008, a research experiment was conducted for the MEPS at the request of the U.S. Office of Management and Budget (OMB), as sponsored by the Agency for Healthcare Research and Quality and the Centers for Disease Control & Prevention. Incentive payments of $30, $50, and $70 were compared among close to 10,000 households in the MEPS 2008 sample panel. The experiment was carried out in five rounds of data collection and the comparable comparison to PIAAC is for the first round. The MEPS is at a similar burden level as PIAAC. In terms of response rates, in the first round, the two higher incentive payment groups had significantly higher response rates than the $30 payment group. Likewise, there was a simultaneous drop in refusal rates, in which the two higher incentive payments had significantly lower refusal rates than the $30 payment group. In Round 1, the difference between the $50 and $70 was not significant, and our understanding is that OMB approved the $50 incentive payment.

The PIAAC field test included an experiment to evaluate the impact of increasing the incentive amount from $35 (equivalent to 2003 ALL and NAAL incentives when accounting for inflation) to $50 to account for the added burden of a longer interview and assessment than past literacy surveys and the increased complexity of the PIAAC computer-based assessment.

An experiment was included in the field test to evaluate the impact of increasing the incentive amount from $35 (equivalent to 2003 ALL and NAAL incentives when accounting for inflation) to $50 to account for the added burden of a longer interview and assessment (about two hours total) than past literacy surveys and the increased complexity of the PIAAC computer-based assessment. During the field test, 1,510 noninstitutionalized adults 16 to 65 years old were assessed in 22 Primary Sampling Units (PSUs) across the country. This paper provides a report on the results of the incentive experiment, which analyzed the incentive payment impact on refusal rates, level of effort, and item nonresponse rates. At the time of the writing of this paper, the field test and main study data collections have been completed.

2. PIAAC Background

The PIAAC survey is a multi-cycle study that is a collaborative effort between the governments of participating countries, the Organization for Economic Cooperation and Development (OECD), and a consortium of various international organizations, referred to as the PIAAC Consortium, led by the Educational Testing Service (ETS).

The study assessed the following adult skills required in the information age: basic reading skills, reading literacy, numeracy, and problem solving in "technology-rich environments" (the OECD term for 'on or with a computer'). PIAAC also measured the ability of individuals to use computer and web applications to find, gather, and use information, and to communicate with others. The study used a "Job Requirements Approach" to ask employed adults about the types and levels of a number of specific skills used in the workplace. These included not only the use of reading and numeracy skills on the job, but also physical skills (e.g., carrying heavy loads, manual dexterity), people skills (e.g., public speaking, negotiating, working in a team), and information technology skills (e.g., using spreadsheets, writing computer code). It asked about the requirements of the person's main job in terms of the intensity and frequency of the use of such skills. PIAAC also broke new ground by being the first to use computers to administer an international assessment of this kind, though some individuals were given a paper and pencil version of the assessment.

An important element of the value of PIAAC was its collaborative and international nature. All PIAAC countries followed common standards and procedures and used the same virtual machine (VM) software when conducting the survey and assessment. As a result, reliable and comparable measure of adult skills in the adult population (age 16-65) of participating countries will be produced.

3. Sample Design

Discussing the main study sample design first will help to put into context the design for the field test sample. The PIAAC target population consists of non-institutionalized adults who reside in the U.S. (whose usual place of residency is in the country) and are between the ages of 16 and 65 years, inclusive, at the time of the interview. Adults are to

be included regardless of citizenship or language. The main study assessment was comprised of a probability-based nationally representative sample of 5,000 persons. The standard PIAAC design required calculable probabilities of selection at each stage of sampling for the main study. Thus, each person in the target population had a known non-zero probability of selection. A four-stage sample design was employed in which the PSUs were counties or groups of contiguous counties. The second stage units were segments (census blocks or combinations of blocks), the third stage units were dwelling units (DUs), and the fourth stage units involved selecting one or two eligible adults per household. Once dwelling units were selected, a screener interview was conducted to identify the eligible persons within selected households. A sampling algorithm was 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 was completed. Upon completion of the BQ, the selected person answered a small number of core assessment items in two stages. In the first stage, the core items were related to the ability to use the computer. If the respondent passed the computer core assessment, then the respondent was routed to the cognitive core assessment via the computerized module. If the respondent did not pass the computer core assessment, the respondent was routed to the cognitive core assessment via the paper module. If the respondent passed the cognitive core assessment, the respondent was given either the computerized main assessment or the paper main assessment, depending on the results from the stage 1 core assessment. Those who did not pass the core assessment were done with the interview. In total, it took about two hours for a respondent to complete all PIAAC instruments.

For the field test, the workflow was different from the main study due to the need to analyze the differences between those who take the computer assessment and those who would take the paper assessment. To do so, a computer core assessment was conducted, and among those who passed the computer core, a portion was randomly assigned to a computer assessment and a portion was randomly assigned to a paper assessment. The field test sample design was impacted because it was necessary to ensure that there were enough computer literate respondents for the computer/paper assessment comparison. The field test sample design essentially followed closely to the four-stage structure of main study design, however, we note the following exceptions. The field test was a household nonprobability sample with the goal of completing assessments for 1,500 adults ages 16 through 65 years. A four-stage sample design was employed in which the PSUs were purposively selected to make use of existing listings, to increase the chance of passing the stage 1 core assessment on computer skills, and to reduce the need for bilingual interviewers. The second stage units were randomly selected segments (census blocks or combinations of blocks), the third stage units were randomly selected DUs, and the fourth stage units involved the random selection of one or two eligible adults per household. The fourth stage of selection started with a screener interview to identify the eligible persons within selected households. A sampling algorithm was implemented within the CAPI system to select one or two sample persons among those identified to be eligible.

After the completion of six weeks of data collection, the sample monitoring reports predicted sample yield rates higher than the initial estimates. Therefore, in the seventh week, due to the high cost of completing more cases than required, to reduce the total number of completes at the end of data collection, DUs were deselected using a ratio of 1/3 (selecting 2/3) of the interim cases identified by “not home” and “not worked.” There were 358 cases deselected at that time. This impacted the weights that were used for the incentive experiment, as described later.

4. Experiment Design

The incentive experiment was conducted at the segment level (clusters of DUs within PSUs). The experiment was not conducted at the DU level because such designs have an increased chance of introducing error in administering the incentives to the respondents, and because of the risk of spreading information about different incentive amounts in close neighborhoods.

Incentive payments were randomly assigned to each segment. By doing so, each interviewer was assigned both incentive amounts to minimize any interviewer impact. The achieved response rates for each stage of data collection, and the overall response rate for each incentive group are provided below.

5. Experiment Response Rate

The initial overall (unweighted) response rate in the $50 group was 5.2 percentage points higher than the response rate for the $35 group. The screener response rate among DUs in the $50 group was 3.2 percentage points higher than the response rate for the $35 group. The BQ response rate among respondents in the $50 group was 4.0 percentage points higher than the response rate for the $35 group. The assessment response rate was the same for both groups (96.5 percent).

These rates, however, do not take into account for (1) the fact the field test sample was purposefully selected from areas with high computer literacy[1] and (2) the fact that not all persons selected into the sample became aware of the incentive offered to them (even though advanced letters were mailed to all households explaining the incentive).[2] In order to account for the fact the field test sample was selected from areas with high computer literacy, weights were assigned to the sample cases so that the total sample would reflect the population distribution of the United States according to the percentages of the following variables: "less than a high school education," "average earnings below 150 percent of the poverty line," and "Black or Hispanic." In order to account for the fact that not all persons selected into the sample became aware of the incentive offered to them, an "experiment response rate" was calculated using the cases that remain in the experiment once those unaware of the incentive were dropped. The remaining cases consisted of all completes,[3] refusals, and partial complete or breakoffs. Thus the *experiment response rate* = completes / [completes + refusals + partial complete or breakoffs]. Sample cases that were never contacted were excluded from the analysis since the incentive payment did not have any effect on their response status.

[1] The PSUs for the field test was selected as a non-probability sample, chosen with the following goals: Satisfy the demographic requirement of the psychometric testing; and optimize the ICT Core passing rate to achieve 1,300 completed assessments who passed the ICT Core instrument.

[2] Some selected persons were unaware of the incentive amount on account of a language problem, refusal by gatekeeper or another person to inform them, learning/mental disability, reading/writing difficulty, impairments (hearing, blindness/vision, speech), disabilities (physical, other), other unusual circumstances, no contact before maximum number of calls reached temporarily absent, vacant/not DU/under construction, and death.

[3] The category of "complete" cases includes screeners that were completed but did not have a person in the target population (16-65 year olds) in the household.

6. Analysis Results for Refusal Rates

The (weighted) experiment response rate is the appropriate statistic for assessing the incentive experiment; however, to avoid the potential for confusion having two different sets of field test response rates in various documents, it was deemed best to analyze the field test data using the complement of the experiment response rate, referred to as the *refusal rate* and defined as:

refusal rate	=	1 – experiment response rate
	=	[refusals + partial complete or breakoffs]/ [completes + refusals + partial complete or breakoffs]

The refusal rates for the two incentive levels in the field test, after accounting for the field test design, differed as follows:

- The overall weighted refusal rate in the $50 group (28.7 percent) was 6.8 percentage points lower than the weighted refusal rate for the $35 group (35.5 percent).
- The screener weighted refusal rate among DUs in the $50 group was 0.6 percentage points lower than the response rate for the $35 group.
- The BQ weighted refusal rate among respondents in the $50 group was 6.2 percentage points lower than the weighted refusal rate for the $35 group.

The statistical analysis described in the remainder of this report concluded that the difference in the overall refusal rate between the two incentive amounts is significant at the 0.05 level. That is, there was enough evidence to show that the $50 incentive amount had a significantly lower refusal rate when compared to the $35 incentive amount. The screener refusal rate and BQ refusal rate were also tested individually. The difference between the screener refusal rates for the two incentive levels was not significant; however, the difference between the BQ refusal rates for the two incentive levels was significant at the 0.05 level.

Logistic regression modeling was used to test the null hypothesis of no impact on the chance for refusal between $35 and $50 incentives, after controlling for other variables in the model. In addition to measuring the incentive payment effect on refusal rates, the model also estimates effects of other variables on refusal rates. A stepwise logistic regression was processed to select explanatory variables for the model. Explanatory variables, relating to race/ethnicity, education attainment, median income, Metropolitan Statistical Area (MSA) status, and poverty status, were gathered from Census 2000 data at the segment level. Variables for the BQ level analysis also included the person-level variables: age, sex, and race/ethnicity. The stepwise regression helped to address any issues with multicollinearity, which would violate modeling assumptions relating to the independent effects among explanatory variables. Once the set of explanatory variables was selected, a logistic regression model that incorporated clustering effects and weights was processed.

The modeling approach measured the overall impact of incentive payments on refusal rates. To investigate the impact of different levels of incentives on each demographic subgroup (as defined below) individually, simultaneous statistical t-tests were conducted to test the null hypothesis of no difference between refusal rates for the two incentive amounts, by subgroups created based on demographic characteristics of the PSUs in the

sample. The t-tests and the regression analysis were conducted using weights. The weights were set equal to one, except for cases in the Not-at-homes (NH) and Not-worked (NW) strata at the time of deselection of DUs during the data collection period. To account for the deselected cases, the weights for the retained NH and NW cases were set equal to the inverse of the subsampling rate (2/3). In addition, since households are clustered within segments, and segments clustered within PSUs, replicate weights were created for the analysis to capture the clustering effect on variances. The paired jackknife replication approach, also referred to as JK2, was used to facilitate the variance estimation.

All statistical tests were conducted at the 0.05 level of significance to be consistent with the NCES statistical standards. Also, the Bonferroni approach is used to control the level of Type I error when conducting simultaneous multiple comparisons.

6.1 Overall Refusal Rates (Screener and BQ Combined)

The analyses of the overall refusal rates for the screener and BQ stage combined takes into account the cumulative impact of the incentive payment on refusal rates (i.e., refusal at either the screener or BQ stage). The refusal rate is computed using the following definitions for the numerator and denominator of the ratio:

- *Numerator*: Number of selected persons with status as: refusal or incomplete (i.e., partial-completes due to break-offs), and number of selected DUs with status as: refusal or incomplete.
- *Denominator*: The value of the numerator plus the number of selected persons with a completed BQ.

The estimated difference in the overall refusal rates between the $50 and $35 incentive amount is 6.8 percentage points with an associated p-value of 0.018, indicating a statistically significant difference between refusal rates for the two incentive groups.

The probability of a contacted person being a refusal or incomplete is estimated with the following logistic regression model, in which *Y* is a dichotomous variable with a value of 0 if the person is a complete (i.e., completed the BQ), and a value of 1 if the person is a refusal or incomplete. The logistic regression model estimates the probability of the occurrence of Y=1 for case *i* by a function of *k* explanatory variables, as follows:

$$E(Y_i)\frac{\exp(x_i'\boldsymbol{\beta})}{1+\exp(x_i'\boldsymbol{\beta})}$$

Where,

$\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)'$ is a k by 1 column vector of regression coefficients,
$\boldsymbol{x_i'} = (x_{1i}, x_{2i}, \ldots, x_{ki})'$ is the vector of k explanatory variables for case i.

Table 1 contains the results of the logistic regression analysis. The incentive group has a p-value = 0.0006; strong evidence of a statistically significant difference in refusal rates between the two incentive groups. The significant effect implies a lower chance of refusal for the $50 incentive group. In addition, the model also shows that higher segment-level median income indicates a significantly higher chance for refusal, while living in the Midwest or in areas with higher concentrations of non-Hispanic blacks indicates a significantly lower chance for refusal.

Table 1: Logistic Regression Model Parameters and Significance Levels for the Overall Refusal Indicator

Parameter	*Parameter estimate*	*Standard error*	*P-value*
Intercept	-1.35	0.292	0.0001
Incentive group	-0.30	0.076	0.0006
High education	-0.19	0.168	0.2665
Median income in segment	0.03	0.005	0.0000
Midwest	-0.30	0.137	0.0377
Percentage earning less than 150% of the poverty line	-0.94	0.463	0.0548
Percentage non-Hispanic black in segment	-0.82	0.336	0.0227

Source: 2010 Programme for the International Assessment of Adult Competencies Field Test
Note: Median income was divided by 1000.

In addition, simultaneous t-tests were conducted to evaluate the impact of the higher incentive amount on refusal rates for various subgroups individually, based on demographic characteristics of the PSUs. The following area-level subgroups demonstrate statistically significant differences (at the Bonferroni family of statistical testing level of 0.05) in refusal rates between incentive groups (all having lower refusal rates for the $50 incentive amount): PSUs in low poverty areas, high education areas, high black areas, low Hispanic areas, and PSUs in the West, and in MSAs. Although, some of these subgroups demonstrated a significant drop in refusal rates between the $35 and $50 incentive groups, while others did not, the estimated subgroup differences in refusal rates are fairly steady: between 4 and 8 percentage points reduction with the higher incentive level. The results indicate the impact of smaller sample sizes and clustering on the stability of the estimated standard errors for each subgroup.

The following provides the results of the analysis of refusal rates separately for the screener and the BQ stage.

6.2 Screener-Level Refusal Rate Analysis

The refusal rate is computed for the screener with the numerator and denominator defined as follows:

- *Numerator:* Number of refusals or incompletes (i.e., partial-completes due to break-offs).
- *Denominator:* The value of the numerator plus the number of completed screeners (including age eligible or not).

The denominator excludes the following cases for which it is assumed that the incentive payment has no impact: language problem, refusal-gatekeeper, learning/mental disability, impairments (hearing, blindness/vision, speech), disabilities (physical, other), other unusual circumstances, vacant/not DU/under construction, maximum number of calls, temporarily absent.

The difference in refusal rates for the two incentive groups (0.6 percentage points) is not statistically significant.

Table 2 provides the analysis results from the logistic regression model. As shown below, the p-value for the incentive group is 0.0725. Although the p-value is significant at 0.10 level, there is not enough evidence to show a significant incentive group effect on the refusal indicator under the NCES standard significance level of 0.05. The model also shows that there is a significantly higher chance for refusal (at the .05 level) for those with higher segment-level median income, while living in the Midwest indicates a significantly lower chance for refusal.

Table 2: Screener Level Logistic Regression Results on the Refusal Indicator

Parameter	*Parameter estimate*	*Standard error*	*P-value*
Intercept	-2.64	0.393	0.0000
Incentive group	-0.20	0.105	0.0725
Median income for the segment	0.03	0.009	0.0011
Midwest	-0.39	0.137	0.0086
Percentage less than high school attainment in segment	-1.11	0.641	0.0976
Percentage non-Hispanic black in segment	-0.66	0.361	0.0810
MSA status	0.31	0.171	0.0812

Source: 2010 Programme for the International Assessment of Adult Competencies Field Test
Note: Median income was divided by 1000.

Simultaneous t-tests were also conducted to test the differences between incentive payments for each demographic subgroups individually. The sample size is not adequate enough to provide evidence to show significant differences under the NCES standard level of 0.05 for any of the subgroups.

6.3 BQ-Level Refusal Rate Analysis

The refusal rate is computed for the BQ conditional on completing the screener. The numerator and denominator of the ratio are:

- *Numerator*: Number of selected persons with status as: refusals or incompletes (i.e., partial completes due to break-offs).
- *Denominator*: The value of the numerator plus the number of completed BQs.

The denominator excludes the following cases for which it is assumed that the incentive payment has no impact: language problem, reading/writing difficulty, refusal by other person, learning/mental disability, impairments (hearing, blindness/vision, speech), disabilities (physical, other), other unusual circumstances/death, and maximum number of calls.

The estimated difference in the incentive group refusal rates for the BQ is 6.2 percentage points, with an associated p-value of less than 0.001. Thus, the refusal rate for the BQ with the $50 payment is significantly lower than the refusal rate for the BQ with the $35 payment.

The logistic regression analysis results are given in Table 3. The results show a statistically significant difference (p-value = 0.0436) in refusal rates between a $35 and a $50 incentive. This is a more powerful test than the t-test since it controls for all other variables in the model, including variables at the person- level collected from the screener

questionnaire in addition to variables based on area-level percentages described above. The model also shows that there is a significantly higher chance for refusal for those living in the Northeast.

Table 3: Background Questionnaire Level Logistic Regression Results on the Refusal Indicator

Parameter	*Parameter estimate*	*Standard error*	*P-value*
Intercept	-1.53	0.210	0.0000
Incentive group	-0.30	0.138	0.0436
MSA status	-0.32	0.238	0.1866
Hispanic[1]	-0.39	0.315	0.2329
Non-Hispanic black[1]	-0.52	0.371	0.1734
Northeast	0.58	0.157	0.0013
Percentage non-Hispanic black in segment	-0.58	0.383	0.1433

[1] Person-level variables collected from the screener questionnaire.
Source: 2010 Programme for the International Assessment of Adult Competencies Field Test

Simultaneous t-tests were also conducted for the BQ subgroups. In addition to the PSU level subgroups used in the above analysis, the BQ subgroups include the person's age, gender, and race/ethnicity collected from the screener questionnaire. The following subgroups have statistically significant differences (at the Bonferroni family of statistical testing level of 0.05) in refusal rates between incentive groups, all having lower refusal rates for the $50 incentive amount: Non-Hispanic blacks, high poverty areas, low education areas, and PSUs that are in the Northeast area.

7. Analysis Results for the Number of Attempts

Analyses of the impact of incentive payment on the number of attempts were performed at the screener level and the BQ level.

7.1 Screener-Level Number of Attempts Analysis

The screener-level analyses were conducted using the set of DU records comprised of completed screeners (age eligible or not), partial completes/break-offs, and refusals. First, weighted mean numbers of attempts at the screener level and their standard errors were calculated, and the differences between the two incentive groups, overall, and by subgroup were tested. The results indicated that the additional incentive did not have a significant impact on the number of attempts at the screener level.

A multiple regression analysis was also conducted to assess the impact of incentive payment on the number of attempts, controlling for other variables in the model. Since the distribution of the number of attempts was substantially skewed, a linear model was not appropriate for the regression analysis. As an alternative, a transformed regression model was fit where a natural logarithm of the number of attempts at the screener level was taken as the dependent variable. The results of the regression analysis were consistent with those based on the comparison of the weighted sample means.

7.2 BQ-Level Number of Attempts Analysis

The BQ-level analysis was conducted on the set of person records conditioned on complete screeners only. Similar to the screener-level analyses, weighted means for the number of attempts were calculated and tests for the difference between incentive groups

were performed at the BQ level. The results indicated that the group with a higher incentive amount had a smaller mean number of attempts (mean=2.33) than that of the low incentive group (mean=2.47). Such an estimated effect was significant at the .05 level, although the effect was small. The analysis also showed that the effect of the incentive varied by subgroup, suggesting that there were interaction effects between the incentive amount and some demographic and geographic variables.

A multivariate analysis was also conducted by fitting a transformed regression model where the natural logarithm of the BQ-level number of attempts was used as the dependent variable. Table 4 shows the BQ-level regression output. In the table, the regression coefficient estimate for the incentive group is -0.0358, although it is not significant at the .05 level. This estimate indicated that the higher incentive tended to slightly reduce the number of attempts, with other factors being held constant. This result was consistent with that based on the comparison of the weighted sample means.

Table 4: BQ-level Regression Results on the Number of Attempts

Parameter	*Parameter estimate*	*Standard error of estimate*	*Test for h0:parameter=0*	*Prob>|t|*
Intercept	0.8684	0.0538	16.1469	0.0000
Incentive group	−0.0355	0.0188	−1.8877	0.0723
High education	0.0967	0.0283	3.4168	0.0025
Low black	−0.1536	0.0271	−5.6639	0.0000
Northeast	0.3196	0.0513	6.2301	0.0000
MSA status	−0.1095	0.0509	−2.1515	0.0427
Percentage less than high school attainment in segment	0.3704	0.1632	2.2697	0.0334
Percentage non-Hispanic black in segment	−0.1877	0.0658	−2.8523	0.0093
Percentage Hispanic in segment	−0.1896	0.0822	−2.3057	0.0309

8. Item Nonresponse

The item nonresponse analysis was conducted on the set of person records comprised of BQ respondents who were involved in the incentive experiment. The item nonresponse records consisted of item refusals and those who answered “Don’t Know” to the question.

Tabulations on earnings, employment, education, and selected demographic variables found that, except for three earnings variables, item nonresponse was negligible for almost all survey variables. The three earnings variables with a substantial item nonresponse rate (over one percent) are “gross pay,” “additional earnings,” and “total earnings last year.” Weighted item nonresponse rates were calculated for the three earnings variables, and the differences in the rate between the two incentive groups were tested. Because there were valid missing values due to skip patterns, two versions of item nonresponse rates were calculated, one for all records, and the other with the valid skip records excluded.

Table 5 shows the weighted item nonresponse rates for the three earnings variables and the tests of the differences between the two incentive groups. It can be seen from the table that, for all three variables, and for both versions of the rates, the item nonresponse

rates for the higher incentive group are slightly higher than those of the lower incentive group, which is contrary to what was expected. However, none of the differences are significant at the .05 level.

Table 5: Weighted Item Nonresponse Rate for Selected Survey Variables by Incentive Group

	$35 incentive group		*$50 incentive group*		*Test of difference*			
Survey Variable	*Estimate*	*SE*	*Estimate*	*SE*	*Diff*	*SE*	*t VALUE*	*PROB >|T|*
Current work - Earnings - Gross pay								
All records (n=1485)	0.0483	0.0085	0.0660	0.0095	0.0177	0.0142	1.2437	0.2267
Valid skip excluded (n=897)	0.0819	0.0148	0.1078	0.0157	0.0259	0.0240	1.0818	0.2911
Current work - Earnings - Additional								
All records (n=1485)	0.0268	0.0067	0.0318	0.0049	0.0050	0.0084	0.5974	0.5563
Valid skip excluded (n=285)	0.1459	0.0353	0.1617	0.0247	0.0158	0.0421	0.3744	0.7117
Current work - Earnings - Total earnings last year								
All records (n=1485)	0.0274	0.0060	0.0371	0.0070	0.0097	0.0085	1.1358	0.2683
Valid skip excluded (n=153)	0.2485	0.0534	0.3865	0.0570	0.1380	0.0720	1.9162	0.0684

9. Summary

The statistical analyses concluded that the difference in the overall refusal rate between the two incentive amounts is significant at the 0.05 level. That is, there was enough evidence to show that the $50 incentive amount had a significantly lower refusal rate when compared to the $35 incentive amount. The screener refusal rate and BQ refusal rate were also tested individually. The difference between the screener refusal rates for the two incentive levels was not significant; however the difference between the BQ refusal rates for the two incentive levels was significant at the 0.05 level. The statistical

analyses also indicated that increasing the incentive amount from $35 to $50 may slightly reduce the number of attempts at the BQ level. Analyses for both refusal rates and the number of attempts suggest that the higher incentive had a larger effect at the BQ level than at the screener level. The statistical analyses found no significant difference in item nonresponse between the two incentive amounts, which may be attributed to the fact that item nonresponse rate was very low.

References

Mohadjer, L., Berlin, M., Rieger, S., Waksberg, J., Rock, D., Yamamoto, K., Kirsch, I., Kolstad, A. (1997). The role of incentives in literacy survey research, Chapter 10 pp 209-244 in Adult Basic Skills: Innovations in Measurement and Policy Analysis, eds. Tuijnman, Kirsch, and Wagner, Hampton Press, 1997.

Singer, E., Van Hoewyk, J., Gebler, N., Raghunathan, T., & McGonagle, K. (1999). The effect of incentives on response rates in interviewer-mediated surveys. *Journal of Official Statistics, 15* (2), 217-230.