

Appendix G

Design and Analysis of Incentive Experiment

Appendix G Design and Analysis of Incentive Experiment

G.1 Experiment Design

As explained in Statement A, the PIAAC field test included an experiment to evaluate the impact of increasing the incentive amount from \$35 to \$50. This appendix describes this incentive experiment.

The experiment was conducted at the segment level (clusters of dwelling units (DUs) within Primary sampling units (PSUs)) rather than at the DU level because (1) such designs have an increased chance of introducing error in administering the incentives to the respondents, and (2) such designs introduce the risk of spreading information about different incentive amounts in a single neighborhood. Incentive payments were randomly assigned to each segment. The assignment was done systematically by sorting the segments by PSU and geographical location of the segment within the PSU and then alternating the assignment of the \$35 payment and the \$50 payment to the segments. After the assignment, quality checks ensured that each incentive group was balanced in terms of demographic characteristics

(poverty status, educational attainment, percent Black and/or Hispanic, and geographic region).

The following highlights other major aspects of the incentive experiment design.

- Interviewers were given both \$35 and \$50 segments to minimize any interviewer impact on the incentive payment effect.
- Incentive assignment to segments with common boundaries, or in close proximity, followed its natural probability-based assignment. Therefore, there was no special re-allocation of the incentive

groups in order to have the same incentive amount for segments close in proximity.

- The official sample for the incentive experiment were the 3,581 DUs that were released originally after the de-selection process occurred.¹ The released cases from additional reserve sample² were excluded because the resulting nonrespondents may not have been worked fully to meet PIAAC standards.

¹ 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, dwelling units (DUs) were deselected using a ratio of 1/3 (selecting 2/3) of the interim cases identified by DISP_SCR = NH (not home) and

G.2 Method of Analysis

Statistical analyses were conducted to examine differences between the following rates at the two incentive levels:

- Refusal rates for the screener,
- Refusal rates for the BQ,
- Refusal rates for the screener and BQ combined (referred to as “overall” in this document).

NW (not worked). There were 358 cases deselected.

² In the ninth week, a survey control file was prepared for about 150 reserve sample cases, which were released and worked from three select PSUs to ensure that the completed assessment goals were met by the end of the Field Test and to allow us to practice the operations and systems activities required to release additional sample

As explained in Part A, the refusal rate (the complement of the experiment response rate) was selected for the analysis (instead of the unweighted initial response rate). The reason again for not using the unadjusted response rate was that it does not take into account (1) the fact the field test sample was purposefully selected from areas with high computer literacy³ 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

in advance of the Main Study.

³ The PSUs for the Field Test was 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

to all households explaining the incentive).⁴ The reason the refusal rate was adopted instead of the experiment response rate was to avoid the potential for confusion in having two different sets of field test response rates in various documents. The *refusal rate* is defined as:

$$\text{refusal rate} = 1 - \text{experiment response rate}$$

which is to say,

completed assessments who passed the ICT Core instrument.

⁴ 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

$$\text{refusal rate} = \frac{\text{[refusals + partial complete or breakoffs]}}{\text{[completes + refusals + partial complete or breakoffs]}}$$

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

maximum number of calls reached temporarily absent, vacant/not DU/under construction, and death..

to determine the best explanatory variables for the model. Explanatory variables, relating to race/ethnicity, education attainment, median income, Metropolitan Statistical Area status, and poverty status, were gathered from Census 2000 data at the segment level. Variables for the BQ level analysis also included 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 were selected, a logistic

regression model that incorporates clustering effects and weights was processed.

The modeling approach measures 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.

G.2.1 Analysis of Refusal Rates for the combined Screener and BQ stage

The analysis 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 rates⁵ between the \$50 and \$35 incentive amount is 6.8 percent 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.,

⁵ The overall refusal rates were computed using weights that were assigned to sample cases so that the total sample would reflect the population distribution of the United States according to the percentage with less than a high school education, percentage earning below 150 percent of the poverty line, and percent Black or Hispanic.

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:

$$E(Y_i) = \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta})}{1 + \exp(\mathbf{x}'_i \boldsymbol{\beta})}$$

where,

$\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_k)'$ is a k by 1 column vector of regression coefficients,

$\mathbf{x}'_i = (x_{1i}, x_{2i}, \dots, x_{ki})'$ is the vector of k explanatory variables for case i .

Table G-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 lower concentrations of non-Hispanic blacks indicates a significantly lower chance for refusal.

Table G-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. Table G-2 shows the refusal rates for each incentive group, their standard errors, the estimated difference between the refusal rates (and the standard error), and the p-value for the statistical test. The subgroups listed in Table G-2 are defined using data from the 2000 Census for each PSU.

The assignment of PSUs to each subgroup was accomplished in the following way:

- A PSU is classified as *high poverty* if the population percentage earning below 150 percent of the poverty line is greater than 21.4 percent. It is *low poverty* otherwise.
- A PSU is *high education* if the population percentage with no more than a high school education is less than 50.9 percent. It is *low education* otherwise.
- A PSU is *high black* if the population percentage non-Hispanic black is greater than 11.6 percent. It is *low black* otherwise.

- A PSU is *high Hispanic* if the population percentage Hispanic is greater than 7.5 percent. It is *low Hispanic* otherwise.
- Regions are also used as well as Metropolitan Statistical Area (MSA) status.

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.

Table G-2. Overall refusal rates and standard errors by incentive amount, including estimated differences and the p-values

	Sample Size	\$35 Incentive group Estimate	Standard Error	\$50 Incentive group Estimate	Standard Error	Difference	Standard Error	p-value
High poverty	903	36.3%	1.92%	31.4%	3.38%	4.9%	3.14%	0.135

Contacting Households [For Countries With Probability Sample]

6

								4
								0.004
Low poverty	1357	38.5%	1.66%	31.0%	1.97%	7.5%	2.38%	6
High education	1535	39.0%	1.47%	31.7%	1.73%	7.3%	2.10%	2
Low education	725	34.4%	2.42%	30.1%	4.08%	4.4%	3.53%	0.228
High black	1248	35.1%	1.85%	27.8%	2.23%	7.2%	2.62%	2
Low black	1012	40.7%	1.74%	35.4%	2.70%	5.3%	2.94%	0.011
High Hispanic	850	41.3%	2.10%	34.1%	3.50%	7.2%	4.15%	4
								0.086
								6
								0.096
								8

Contacting Households [For Countries With Probability Sample]

6

Low Hispanic	1410	35.4%	1.61%	29.5%	1.96%	6.0%	1.91%	0.0049
Northeast	273	49.5%	1.16%	41.3%	6.15%	8.2%	5.02%	0.1190
Midwest	782	34.3%	2.39%	28.9%	3.44%	5.4%	2.88%	0.0726
South	756	35.6%	2.25%	28.9%	2.57%	6.6%	4.09%	0.1193
West	449	39.5%	2.79%	33.3%	2.22%	6.2%	2.28%	0.0128
NonMSA	339	32.2%	4.12%	30.1%	3.49%	2.1%	4.18%	0.6190
MSA	1921	38.5%	1.35%	31.4%	2.01%	7.2%	2.16%	0.0031

Source: 2010 Programme for the International Assessment of Adult Competencies Field Test
 Note: n = 2260

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

G.2.2 Screener

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 rates⁶ for the two incentive groups (0.6 percent) is not statistically significant.

Table G-3 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-

⁶ The overall screener refusal rates were computed using weights that were assigned to sample cases so that the total sample would reflect the population distribution of the U.S. according to the percentage with less than a high school education, percentage below the 150 poverty line, and percent Black or Hispanic.

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 G-3. Screener level logistic regression results on the refusal indicator

Parameter	Parameter	Standard Error	P-value
	Estimate		
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.

Table G-4 provides the results of simultaneous t-tests for testing 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.

Table G-4. Screener refusal rates and standard errors by incentive group, and estimated difference, with p-values

Subgroup	Sample size	\$35 Incentive group		\$50 Incentive group		Difference	Standard Error	p-value
		Estimate	Standard Error	Estimate	Standard Error			
High poverty	1059	21.2%	1.72%	21.1%	2.87%	0.1%	2.65%	0.9654
Low poverty	1610	25.4%	2.07%	20.1%	1.69%	5.4%	3.04%	0.0906
High education	1756	27.1%	1.58%	22.0%	1.67%	5.1%	2.45%	0.0505
Low education	913	16.9%	2.45%	17.7%	3.15%	-0.8%	3.27%	0.8098
High black	1482	23.0%	2.11%	18.1%	1.75%	5.0%	2.78%	0.0877
Low black	1187	24.7%	1.95%	23.6%	2.58%	1.1%	3.40%	0.747

Contacting Households [For Countries With Probability Sample]

6

High Hispanic	980	27.2%	3.34%	23.7%	2.84%	3.5%	4.59%	7 0.458
Low Hispanic	1689	21.9%	1.26%	18.7%	1.75%	3.2%	2.22%	0 0.168
Northeast	299	29.6%	2.56%	31.2%	7.32%	-1.6%	6.55%	0.812 4
Midwest	983	19.7%	1.80%	16.2%	2.56%	3.5%	3.49%	0.330 9
South	882	24.0%	3.34%	20.2%	1.88%	3.8%	3.91%	0.338 5
West	505	27.7%	2.54%	23.3%	2.28%	4.4%	2.73%	0.125 3
NonMSA	445	14.7%	2.63%	14.2%	2.73%	0.5%	4.15%	0.899 8

Contacting Households [For Countries With Probability Sample]
6

MSA	2224	25.5%	1.63%	21.8%	1.78%	3.7%	2.45%	0.145 4
-----	------	-------	-------	-------	-------	------	-------	------------

Source: 2010 Programme for the International Assessment of Adult Competencies Field Test

Note: n = 2,669.

Note: A PSU is high poverty if the population percentage below the 150 percent of the poverty line is greater than 21.4%. It is low poverty otherwise. A PSU is high education if the population percentage with high school education or less is less than 50.9%. It is low education otherwise. A PSU is high black if the population percentage non-Hispanic black is greater than 11.6%. It is low black otherwise. A PSU is high Hispanic if the population percentage Hispanic is less than 7.5%. It is low Hispanic otherwise. MSA = Metropolitan Statistical Area.

G.2.3 Background Questionnaire

The refusal rate is computed for the Background Questionnaire (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 rates⁷ for the BQ is 6.2 percent, with an associated p-value of less than 0.001. Thus, the refusal

⁷ The overall BQ refusal rates were computed using weights that were assigned to sample cases so that the total sample would reflect the population distribution of the U.S. according to the percentage with less than a high school education, percentage below the 150 poverty line, and percent Black or Hispanic.

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 G-5. The results show a statistically significant difference ($p\text{-value} = 0.0436$) 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.

**Contacting Households [For Countries With Probability
Sample]** **6**

Table G-5. 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

Table G-6 includes the results of the simultaneous t-tests for the BQ subgroups. In addition to the PSU level subgroups used in the above analysis, the BQ subgroups include 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.

Table G-6. Background Questionnaire refusal rates and standard errors by incentive group, and estimated difference and the p-values

Contacting Households [For Countries With Probability Sample]

6

	Sample Size	\$35 Incentive group		\$50 Incentive group		Difference	Standard Error	p-value
		Estimate	Standard Error	Estimate	Standard Error			
Screener variables								
16-25 years old	326	12.1%	2.28%	8.1%	2.40%	4.1%	3.34%	0.2341
26-35 years old	333	9.4%	2.69%	9.5%	2.21%	-0.1%	2.89%	0.9627
36-55 years old	695	15.2%	2.41%	10.6%	1.51%	4.6%	2.75%	0.1080
56-65 years old	338	11.6%	2.36%	10.8%	2.11%	0.8%	3.18%	0.7931
Male	788	12.8%	1.62%	10.0%	1.87%	2.8%	1.92%	0.1666
Female	904	12.7%	1.71%	9.9%	1.58%	2.8%	1.92%	0.1591
Hispanic	158	8.0%	3.13%	9.0%	3.37%	-1.0%	4.26%	0.8166
Non-Hispanic black	366	9.3%	2.50%	1.9%	1.04%	7.4%	2.33%	0.0044
Non-Hispanic	1168	14.5%	1.95%	12.3%	1.25%	2.1%	2.06%	0.3136

Contacting Households [For Countries With Probability Sample]
6

other								
Area variables								
High poverty	689	15.0%	1.60%	9.3%	1.49%	5.7%	1.59%	0.0017
Low poverty	1003	11.0%	1.92%	10.3%	1.34%	0.7%	1.91%	0.7206
High education	1120	10.9%	1.65%	9.6%	1.28%	1.3%	1.76%	0.4822
Low education	572	16.4%	2.19%	10.6%	1.69%	5.8%	2.00%	0.0083
High black	954	10.0%	1.87%	8.6%	1.51%	1.4%	1.84%	0.4619
Low black	738	16.0%	1.90%	11.7%	1.15%	4.3%	2.05%	0.0467 ^a
High Hispanic	612	13.7%	2.37%	10.2%	2.04%	3.6%	1.72%	0.0514
Low Hispanic	1080	12.1%	1.67%	9.8%	1.10%	2.4%	1.89%	0.2264
Northeast	187	25.0%	1.46%	11.7%	1.60%	13.3%	2.96%	0.0002
Midwest	609	12.0%	2.84%	11.6%	1.78%	0.4%	3.06%	0.8960
South	568	10.0%	2.26%	7.4%	1.94%	2.7%	1.31%	0.0542

Contacting Households [For Countries With Probability Sample]
6

West	328	11.4%	2.98%	10.3%	1.17%	1.1%	2.76%	0.6954
NonMSA	277	15.8%	4.23%	14.8%	1.19%	1.0%	3.36%	0.7701
MSA	1415	12.1%	1.34%	8.9%	1.17%	3.2%	1.43%	0.0368 ^a

^a Not a significant difference for the Bonferroni multiple comparisons family of tests at the overall $\alpha = 0.05$ level of significance.

Source: 2010 Programme for the International Assessment of Adult Competencies Field Test

Note: n= 1692