An Experiment in Monetary Incentives

M. Berlin, L. Mohadjer, J. Waksberg, A. Kolstad, I. Kirsch, D. Rock, and K. Yamamoto

1. Introduction

The National Adult Literacy Survey (NALS) is a household survey designed to measure and report on the nature and extent of literacy problems facing the population of the United States. Some 15,000 adults aged 16 and older (plus an additional 12,000 adults from the State Adult Literacy Surveys) responded to a one-hour survey that consisted of a 15-minute background questionnaire and a 45-minute literacy exercise assessing their prose, document, and quantitative literacy skills.

In accordance with the contract between the Educational Testing Service in cooperation with Westat, Inc., and the U. S. Department of Education, a field test of some 2,000 adults aged 16 and older was conducted in a sample of 16 Primary Sampling Units (PSUs) drawn from the contiguous 48 states. One of the purposes of the field test was to evaluate the impact of incentives on response rates, performance, and survey costs.

After presenting a brief review of the literature on incentive experiments (Section 2), this paper provides a detailed description of the NALS field-test design (Section 3), along with analyses of the field-test results (Section 4).

2. Literature Review

The widespread use of survey research today as a means of gathering information makes its cost a matter of considerable interest (Reingen & Kernan, 1977). Because much of a survey's cost goes into achieving high response rates, concerned researchers have explored a variety of response inducement techniques over the past 25 years. In one extensive effort to determine which methods of response rate improvement are most effective, Kanuk & Berenson (1975) examined over 75 articles that addressed increasing mail survey response rates. They found that follow-up contact and the use of monetary incentives were the only two methodological procedures that had any empirical impact on response rates.

Much of the research has found that the use of monetary incentives does increase response rates in surveys. In support of this finding, Armstrong (1975) reviewed 18 studies from 14 different researchers and concluded that not only do monetary incentives in mail surveys yield large increases in response rates but that the larger the incentive, the greater the increase in response rate.

In a review of literature focusing on improving survey response rates, Baxter, et al (1984) discussed the theory that most respondents need a reason for taking the time to participate in a survey. Baxter's study indicates that for all respondents there is some small amount of money which functions as a symbolic reward that generates participation. They note, however, that part of the inconsistency in the research on incentives and response rates is that the financial standing of respondents determines the range of amounts that serve as token rewards.

Several reports on personal interview studies and the effect of monetary incentives on response rates are consistent with the mail survey results (Gunn & Rhodes, 1981).

The only study that has not shown improvement on response rate with incentives is a 1972-73 household survey conducted by the Census Bureau for the Bureau of Labor Statistics (Walsh, 1977). However, there were design and operational problems, and the Census Bureau noted that "results might have been different if the experiment had been conducted after resolution of certain of these problems."

3. The NALS Field-Test Design

The NALS field test was conducted in a sample of 16 Primary Sampling Units (PSUs), consisting of counties or groups of counties representing the contiguous 48 states. The PSUs were selected in such a way that satisfied a Latin Square design based on key variables thought to be related to response rates. The variables included region of the country, urbanicity, race/ethnicity, and average income/education level of persons residing in the selected PSUs. On average, about 21 segments (consisting of census blocks) were selected within each PSU, with about 8 households selected in each segment.

Incentive levels of $0, $20, and $35 to be evaluated in the field test were agreed upon by NCES and OMB. Incentives were randomly assigned to segments so that each incentive group had about the same number of cases and there were about the same number of incentive groups represented at the PSU and census region levels.

Approximately two interviewers were assigned to each of the PSUs to complete the field work. Each of 38 interviewers was assigned an approximately equal number of segments in the three payment groups. Interviewers were instructed to introduce the incentive after the household composition was determined and the eligible respondent(s) selected. Respondents were paid by check at the completion of the exercise booklet. Only respondents who completed the background questionnaire and agreed to take the exercise booklet were given the incentive check. In order to compare the costs of the three incentive levels (as well as response rates and other measures of quality of the results), interviewers were required to record time spent and expenses incurred by segment each week.

4. Analyses of the NALS Field Test

The analyses of the field-test data showed that payments of incentives significantly improved the outcomes of the survey. The major areas showing significant improvement by providing incentives to the respondents were the following:

- *Response Rates.* There were significant increases in response rates for the background questionnaire and exercise booklets (incentives were not introduced with the screener questionnaire) as a result of offering incentives to the respondents. Given the sample size, no statistically significant difference in response rates was detected between the $20 and $35 incentive groups.
- *Representation of the Target Population.* Analyses showed that the incentives were most effective in improving response rates for people with low educational attainment and minority populations who are frequently underrepresented in national household surveys (a discussion of the issues related to undercoverage in household surveys is provided later). As a result, the incentives provided a better representation of the general population and improved the representation of subgroups, such as the Black and Hispanic populations, that are of special interest to NALS.
- *Relationship between Incentive Levels, Self-Selection, and Performance.* The accuracy of the estimates of adults' literacy proficiencies depends on gaining the cooperation of the majority of eligible respondents to take the literacy exercise. A significant number of refusals within any one of the three incentive levels will not bias the results only if the refusals do not differ in any relevant or systematic way from the respondents who complete the exercises in the remaining incentive levels. Results of the field test show that if no incentive payment is offered to eligible respondents, the main assessment will **substantially increase bias in estimates** of the population's literacy level. This increase in the bias is likely to result from self-selection factors occurring in the $0 incentive group that lead to a nonrepresentative sample. If no incentive is used in the main assessment, these self-selection factors will result in an overestimate of the literacy levels in the United States for both the total population and major subgroups. Possible remedies would be the use of post-stratification procedures. It is our opinion, however, that these procedures would not completely eliminate the apparent bias.
- *Survey Costs.* A cost analysis of the field-test experiment showed a reduction in interviewing costs when incentives were given to the respondents. The cost per completed interview (including the cost of the incentive) for respondents in the $20 incentive group was lower than in the $0 and $35 incentive groups.

The remainder of this paper provides a more in-depth discussion of the results of the field test with respect to the impact of incentives. The response rates for different incentive levels are presented in Section 4.1. Section 4.2 contains a discussion of the improvements in the representation of the target population when incentives are paid to respondents. The relationship between incentive levels, self-selection, and performance is presented in Section 4.3. The effect on costs is examined in Section 4.4. Finally, Section 5 presents the summary and conclusions of the analyses.

4.1 Response Rates

The survey included three instruments: screener, background questionnaire, and exercise booklet. The screener collected household level data for the purpose of selecting an eligible respondent. If the household had been assigned an incentive, the selected respondent was told that an incentive would be paid upon the completion of the survey instruments. The background questionnaire collected information in six areas: demographic data, language background, education, political and social participation, labor force participation, and literacy activities. The exercise booklet consisted of three 15-minute sections of prose, document and quantitative tasks.

There were 2,774 households in the sample. Out of these 2,774 households, 336 (12.1 percent) units were found either to be vacant or not satisfying the definitions of a dwelling unit at the time of screening. As a result, 2,438 households were found to be eligible, of which 2,155 (88.4 percent) completed the screener. Out of 283 nonresponding households, 152 (6.2 percent) refused to participate in the study, and 131 (5.4 percent) did not complete the screener for other reasons.

In households with completed screeners, one person was randomly selected if there were three or fewer eligible people in the households. Two persons were selected in households with four or more eligible people. After the respondent(s) was selected, the background questionnaire was administered.

A total of 2,288 eligible respondents were selected for the background interview; of that total, 412 did not complete the background questionnaire. Of the 412 who did not complete the background questionnaire, 258 (62.6 percent) were refusals, 39 (9.5 percent) had some type of mental or physical disability, and the remainder were nonrespondents for other reasons (such as broken appointments, language barrier, and so forth).

The response rate analyses indicated a statistically significant improvement in the background questionnaire response rates for respondents given incentives. The response rate increased by about 4 percent when an incentive of $20 was paid to the respondents. There were no statistically significant improvements in response rate when the incentive was increased from $20 to $35.

All 1,876 persons who completed the background questionnaire were asked to take an

exercise. Among the different incentive levels, the same pattern of response rates as observed for the background questionnaire was observed for the exercise booklet. The analysis showed that there was a significant improvement (5 percent) in exercise response rates for respondents given incentives. There were, however, no statistically significant differences between the $20 and $35 incentives. When the $20 and $35 incentive level groups were combined, there was a net gain of 9 percent over the $0 incentive group in the combined response rate for the background questionnaire and the literacy exercise. The following is a summary of response rates by incentive level.

Summary of response rates by incentive levels

Incentive	$0	$20	$35
Screener	87.4%	87.7%	90.0%
Background	78.6%	82.7%	84.4%
Exercise	92.8%	97.9%	98.0%
Overall	63.75%	71.0%	74.4%

Almost all of the nonrespondents were persons who refused to take the literacy exercises. The proportion of those who refused to do the exercise booklet was significantly higher for the $0 incentive group than for the $20 or $35 incentive groups.

4.2 Representation of the Target Population

When a monetary incentive was paid, a disproportionate share of the increase in response rates occurred in subgroups of the population that are of special interest to literacy studies, that is, minorities and undereducated persons. These two groups currently account for over 20 percent of the United States population.

It is particularly important to achieve good response rates for minorities so as to compensate partially for the undercount that almost always occurs in household surveys. The evaluations carried out of the decennial censuses indicate that there has been a recurrent undercount of about two to three percent of the population. Furthermore, since at least 1950, coverage of the Black population in censuses has been lower than that of the White population, with the proportion of uncounted Black males higher than other major demographic subgroups. Recent studies of undercounts have shown that the coverage for the Hispanic population may even be a little lower than that for the Black population. Sample surveys usually do not even do as well as censuses. Consequently, low response rates resulting from not offering an incentive would intensify even further the potential coverage bias.

There was a statistically significant increase of about nine percent in the background questionnaire response rate for the minority population (Black and Hispanic populations combined) as a result of providing $20 incentive. The exercise response rates for the minority populations also increased by about 10 percent as a result of offering significantly $20 incentives. This increase in response rate was also statistically significant. Thus, the overall effect of the $20 incentive was to add about 20 percent to the response rate for minorities. Given the sample sizes, there were no statistically significant differences in response rates between the $20 and $35 incentives.

Persons with less educational attainment made up an important subgroup to study from a policy perspective. The experiment showed a large increase in exercise response rates in the $20 and $35 incentive groups for persons with no high school diploma. The same pattern exists for those with only a high school diploma. There was a significant increase of about 8 percent in the exercise response rate for persons both with no high school diploma and with only a high school diploma when a $20 incentive was paid to the respondents. The increase in response rates for persons with some college or a college degree was not found to be significant.

Finally, the incentive experiment showed a significant increase in response rates for persons aged 16 to 64 when incentives were paid to the respondents. There was an increase of about 8 percent in background questionnaire response rate for persons aged 16 to 64 when $20 incentives were paid and an increase of about 4 percent was observed in exercise response rate. Both the questionnaire and exercise response rates increases were proven to be statistically significant for the 16-to-64-year old population.

4.3 Relationship between Incentive Level, Self-Selection, and Performance

This section evaluates the impact of incentives on the demonstrated distribution of proficiency scores. Through examination of the distributions of estimated literacy proficiency scores, the field-test data provide a means for discovering whether or not the groups agreeing to complete the literacy tasks represent individuals of similar demonstrated literacy proficiency. That is, is there evidence that individuals of higher (or lower) literacy proficiency are more likely to participate under a given incentive condition?

Before discussing the results of the analyses, it may be helpful to describe briefly the assessment booklets (the combination of literacy or cognitive tasks) and the literacy score estimates derived from the field test. Some 100 literacy tasks were developed and assembled into nine discrete blocks, each of which was expected to require 15 minutes of administration time. Each block contained approximately the same number of questions contributing to each of the three literacy scales. The nine blocks of tasks were then assembled into nine booklets, each requiring a total of approximately 45 minutes of administration time. The booklets were configured so that the same three blocks appeared together in three booklets, with each block falling in each position — that is, as the first, middle,

and last block of a booklet. The item response theory (IRT) scaling procedure employed to estimate literacy proficiency scores allowed one to put all scores on a scale, even though groups of individuals took different sets of tasks (Lord 1980). The IRT scale defined for the field test had a mean of 50 and a standard deviation of 10 for the proficiency distribution of the document literacy scales. One key question is whether or not the individuals who did cooperate under the $0 incentive level are different with respect to literacy proficiency from those who cooperated under the other two incentive levels. To address this question, a two-way analysis of variance was conducted where the independent variables were test booklets (three groups of booklets) and three levels of incentives — $0, $20, and $35. The dependent variables were item response theory (IRT) scale scores. The analysis of variance showed that incentive level had a significant impact on performance. The next important question investigate dealt with the nature of the significant impact that incentive level had on literacy performance.

Table 1 presents literacy performance on the, document, literacy scale by incentive levels, as well as by selected demographics crossed with incentive levels. These cross-classifications indicate that the significant analysis of variance results can be accounted for by performance level differences between the $0 incentive group and the remaining two incentive levels. The total mean literacy scores for the $0 incentive level are significantly higher than the corresponding means for the $20 and $35 incentive levels for all three scales. There is no significant difference between the total mean literacy scores for the $20 and $35 incentive levels on any of the three literacy scales. Further inspection of the data indicates that the performance level difference in favor of the $0 incentive individuals is about 20 percent of a standard deviation on each of the three scales. Typically, in the education evaluation literature, a difference in group mean performance of greater than one-tenth of a standard deviation — "a small but nontrivial difference" — is judged to be of practical educational significance (Cohen, 1988).

Table 1. Proficiency means and standard deviations by incentive and scale

	Total		$0		$20		$35	
	N	Mean	N	Mean	N	Mean	N	Mean
Total	1707	50.0	495	50.8	563	49.6	649	49.6
Gender								
Male	775	50.4	225	51.4	271	49.9	279	50.1
Female	917	49.6	264	50.4	290	49.3	363	49.1
Ethnicity								
White	1370	51.2	405	51.7	441	50.9	524	51.0
Black	171	42.7	39	44.4	60	43.6	72	41.0
Hispanic	124	46.4	43	49.2	45	45.5	36	44.1
Asian	26	52.2	5	47.6	11	52.6	10	54.2
Amer Ind	11	46.7	1	30.6	4	44.9	6	50.6
Other	3	49.2	2	52.4	1	42.8		0.0
Education								
In H S	68	50.1	20	49.5	21	50.8	27	50.1
< H S	94	36.6	20	39.2	30	36.3	44	35.6
Some HS	180	43.4	47	43.9	73	43.0	60	43.4
GED/HSEQ	54	46.5	11	47.9	22	45.9	21	46.5
HS Grad	405	48.0	102	48.5	147	47.6	156	48.1
Post Sec	438	52.0	137	52.2	131	51.9	170	51.9
Collgrad	437	56.2	143	56.1	133	56.9	161	55.7
Don't know	3	35.5	1	32.2		0.0	2	37.1
No Ed Us	18	39.3	10	41.8	2	46.2	6	32.8
Age								
16 to 20	156	51.6	44	51.4	48	50.3	64	52.7
21 to 25	153	53.6	42	54.2	45	52.9	66	53.7
26 to 31	211	53.4	48	54.2	82	52.3	81	53.9
32 to 45	544	52.5	155	53.8	202	52.2	187	51.8
46 to 65	356	48.2	103	49.1	106	47.9	147	47.9
65+	287	41.9	103	45.0	80	40.7	104	39.8
Incentive level								
$0	495	50.8	495	50.8	0	0.0	0	0.0
$20	563	49.6	0	0.0	563	49.6	0	0.0
$35	649	49.6	0	0.0	0	0.0	649	49.6

The apparent self-selection in the direction of higher scoring individuals in the $0 incentive group is further evidenced by the fact that $0 incentive group has a disproportionately greater percentage of individuals having some college or a college degree.

Similarly, respondents in the $0 incentive group are proportionately more likely to come from households characterized by relatively high income levels. Those individuals who are 65 years and older and were in the $0 incentive group scored significantly higher than did their counterparts in the $20 and $35 incentive groups.

While there were too few 65-and-over individuals in each of the incentive levels to look for possible bias patterns due to differential self-selection, the significantly higher performance of the $0 elderly group suggested a positive selection bias similar to that found in other $0 incentive groups. As indicated above, this pattern of higher performance in the $0 incentive group was a consistent finding not only for the elderly, but for virtually all other subpopulations on all three literacy scales. However, the incentive level difference was quite small among well educated subpopulations, such as post secondary and college graduate.

Given the mean score differences by incentive level for the entire field-test sample, it is apparent that if no payment incentive was offered to eligible respondents the results of the main assessment are likely to be a nonrepresentative sample that will overestimate the level of literacy in the United States. Moreover, the overestimation is likely to be an even greater problem in the results for a number of subgroups of interest. For example, the $0 incentive condition can be expected to yield skewed distributions for Hispanic adults; for individuals with 0 to 8 years of education and high school and college graduates; for those over the age of 26 and, in particular, for senior citizens and for individuals at all income levels, except possibly the $5,000 to $9,000 level. One possible remedy for this apparent bias would be the use of post-stratification procedures.

Another question that is addressed through the field-test data is the impact of incentive level on motivation to perform. Table 2 presents the results of a two-way analysis of variance booklet groups by incentives, with proportion of items attempted as the dependent variable. To the extent that the number of items attempted is a proxy for motivation, this analysis attempts to evaluate the direct impact of incentive level on motivation to perform after having made the decision to take the literacy exercises. As shown in Table 2, there is no significant relationship between incentive level and proportion of items attempted. It appears from the data that response bias came from self-selection factors that affected whether or not an individual chooses to participate in the assessment. Once an individual agreed to participate, level of incentive paid to a respondent did not seem to have an impact on the individual's motivation to perform as measured by the number of tasks attempted.

Furthermore, the partition analysis was used to separate the differences of total sample means of three incentive levels into the differences in the proportion of subpopulations or the differences in the means of subpopulations. The proportion of the differences attributable to the difference in the subpopulation proportion is identical to the amount of the correction possible to reproduce with incentive level results by applying poststratified weights to the without incentive group results. It was found that the educational level had the highest proportional correction of 40%. However, 60% of the difference could not be corrected. Hence, based on the field test analysis, we concluded that the poststratified weights alone would not completely eliminate the bias. It was crucial to attain the highest participation rate as possible to minimize the bias and maximize the ability to generalize the results.

Table 2. Results of 2-Way ANOVA on proportion of items attempted (Booklet group by incentive)

Source	SS	NDF	MS	F	P
Total	1497.7658	1707			
Mean	1392.6938	1	1392.6938	29296.5878	0.0000
Books	0.1814	2	0.0907	1.9079	0.1487
Incentives	0.1392	2	0.0696	1.4637	0.2317
B*I	0.1140	4	0.0285	0.5997	0.6629
Error	80.7191	1698	0.0475		

4.4 Survey Costs

One final but important component in evaluating financial incentives was to review the impact on survey costs. Thus, record keeping procedures were implemented to allow the analysis of time and expense data by level of incentive payment.

In Table 3, we show the cost per completed assessment in each of the three incentive groups. The cost per completed assessment only includes interviewer wages and expenses (mileage, telephone tolls, etc.). The costs within an incentive group were divided by the number of completed assessments within the incentive group to calculate the average cost per completed assessment.

Table 3. Survey costs and level of effort by incentive

	Incentive Group		
	$0	$20	$35
Average hours per completed assessment	8.4	6.8	6.4
Average # of contacts per completed assessment	6.7	5.3	5.0
Average interviewer costs	$130.42	$98.97	$94.24
Average interviewer costs + Cost of incentive	$130.42	$118.97	$129.24

The field-test experiment indicated that the cost of interviewing and conducting the assessment was reduced when $20 incentives were paid to the respondents, even when the incentive is added to the interviewer costs. This is primarily due to the fact that the incentive reduced the number of contacts (callbacks) the interviewer had to make to complete an assessment. The average number of contacts per completed assessment decreased as the incentive amount increased. When the $20 incentive was added to the interviewer costs, the net cost to the survey was $118.97 as compared with $130.42 when no incentives were paid and $129.24 when the $35 incentive was paid.

5. Summary and Conclusions

The results of this study are relatively straightforward. The analyses of the incentive experiment indicate that a $20 incentive significantly increased the response rate, especially for the subgroups with low levels of literacy. Increasing the response rates for groups that are less likely to participate in the survey will improve the distribution of the sample and the representation of the target population. Furthermore, if no incentive is offered, the likelihood of getting a biased estimate of the population's literacy level is increased. This likelihood is due to the greater tendency for self-selection factors to occur in the $0 incentive group that, in turn, can lead to a nonrepresentative sample. More specifically, under the $0 incentive condition, there is a tendency for the more educated and more wealthy and, thus, more proficient individuals to agree to take the literacy tasks. In other words, those with less educational attainment are less likely to take the literacy exercises when there is no monetary incentive offered. Possible remedies for this apparent bias would be the use of poststratification procedures. It is our opinion, however, that these procedures would not completely eliminate the apparent bias. Having once made the decision to take the literacy exercises, however, the incentive level seems to have little or no effect on an individual's motivation to do his or her best as measured by the number of tasks attempted.

One final consideration with respect to the use of monetary incentives is the total cost of collecting data. The field-test results indicate that the cost of conducting the assessment is less for the $20 incentive group as compared with the costs for the $0 and $35 incentive groups. The net cost per completed interview in the $20 condition is $118.97; for the $0 incentive condition it is $130.42. This reduced cost is due primarily to the fact that the incentive reduced the number of callbacks the interviewer had to make to complete a given survey.

REFERENCES

Armstrong, J. (1975). Monetary Incentives in Mail Surveys. *Public Opinion Quarterly*, Vol. 39, 111-116.

Baxter S., Nieva V. F. and Waksberg, M. (April 1984). Improving survey response rates: A review of the literature. Rockville, MD: Westat, Inc.

Berry, H. and Kanouse, D. E. (1987). Physician Response to a Mailed Survey. *Public Opinion Quarterly*, Vol. 541,102-114.

Cohen, J. (1988). *Statistical power analyses for the behavioral sciences*. New York: Academic Press.

Duncan, W. J. (1979). Mail Questionnaires in Survey Research: A Review of Response Inducement Techniques. *Journal of Management*, Vol. 5, No. 1, 39-55.

Gunn, J. and Rhodes, I. N. (1981). Physician Response Rates to a Telephone Survey: Effects of Monetary Incentive Level. *Public Opinion Quarterly*, Vol. 45, 109-115.

Kanuk, L. and Berenson, C. (1975). Mail Surveys and Response Rates: A Literature Review. *Journal of Marketing Research*, Vol. XII, 440-453.

Linsky, A. S. (1975). Stimulating Responses to Mail Questionnaires: A Review. *Public Opinion Quarterly*, Vol. 39, 82-101.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.

Reingen, P. H. and Kernan, J. B. (August 1977). Compliance with an Interview Request: A Foot-in-the-Door, Self-Perception Interpretation. *Journal of Marketing Research*, Vol. XIV, 365-369.

Walsh, T. C. (1977) Selected results from the 1972-73 Diary Surveys. *Journal of Marketing Research*, Vol. XIV, 344-352.