INTERVIEWER TRAINING TO INCREASE SURVEY PARTICIPATION[1]

Eileen M. O'Brien and Thomas S. Mayer, U.S. Census Bureau
Robert M. Groves, University of Michigan and Joint Program in Survey Methodology, USA
Grace E. O'Neill, Joint Program in Survey Methodology, USA

1. Introduction

The interaction between interviewer and respondent up to the point of the start or termination of the interview is generally short, between 1-5 minutes for face-to-face interviews (Groves and Couper 1998), and even shorter for telephone interviews (Oksenberg, Coleman, and Cannell 1986). Two behaviors of novice interviewers are believed to generate "soft refusals" in these first interactions: pressing the respondent to make a decision too quickly (Groves, Cialdini, and Couper 1992), and ineffectively addressing a specific respondent concern (Groves and Couper 1998). Furthermore, inexperienced interviewers often interpret respondent concerns expressed as questions as signaling reluctance. Questions, however, often lead to complete interviews for a number of reasons: respondents are attending with sufficient interest to pose a question; in inviting a response (by posing a question), respondents are permitting the conversation to continue; and the question provides interviewers with more information to tailor the conversation to that particular respondent. Thus, Groves and Couper suggest refusal aversion training should help interviewers better identify a respondent's main concern, and quickly deliver a brief, situation-appropriate response to that concern.

The theoretical basis Groves and Couper (1998) provide suggests refusal aversion training should include *tailoring* (adapting to specific characteristics of a household) and *maintaining interaction* (continuing the conversation with the respondent). Experienced interviewers are more adept at both. An independent review of a Census Bureau training program suggests that these skills tend to be absent in existing training regimens (Doughty et al., 2000). One focus of this training is to build self-confidence in the difficult job responsibilities described by Doughty et al.

The current study expands on the Groves and McGonagle (2001) refusal aversion training design derived from Groves and Couper's (1998) theory of survey participation by implementing it in a face-to-face household rather than telephone survey environment. In-person interactions afford interviewers many more respondent cues regarding the survey request, (e.g., their household, their verbal responses, their non-verbal responses, etc.).

This paper reports results of a refusal aversion training experiment in a national household health survey. Results suggest that increases in participation rates can also occur in a face-to-face interview environment. In addition, the model to predict future success in obtaining cooperation based on exposure to the experimental training suggested by Groves and McGonagle (2001) is retested. Finally, trainee scores obtained through a revised evaluation methodology are added to measure the relationship between success in training and performance in the field.

2. Method

2.1. Survey and Participants

Participants in this study included 40 Census Bureau field representatives (interviewers) working the National Health Interview Survey (NHIS). The NHIS uses a multi-stage area probability design, oversampling for African American and Hispanic persons. Weekly samples are statistically representative of the non-institutionalized civilian population.

2.2 Materials

Focus Groups. Three focus groups were conducted in order to develop the training materials for the refusal aversion training. The first two were conducted to obtain verbatim verbal concerns most commonly heard by interviewers from respondents. Participants were also asked to suggest environmental cues they observe while approaching a household that may signal reluctance and what non-verbal behaviors a reluctant respondent may exhibit. Researchers grouped the verbal concerns into ten main themes: (a) Purpose; (b) Time (bad timing, burden, burnout); (c) Privacy; (d) Confidentiality; (e) Government; (f) Census; (g) "Why me?"; (h) Voluntary Survey; (i) Consent Form; and (j) Lack of Interest.

The third focus group was conducted in person with very experienced successful interviewers. Their role was to provide effective verbatim interviewer responses, strategies and behaviors for addressing the respondent concerns identified by the first two focus groups.

Refusal Aversion Handbook. The essential building block of the training is the Five Basic Steps to Encouraging Survey Response: (a) Prepare for the visit; (b) Engage in active listening; (c) Diagnose the main concern; (d) Quickly identify a situation-appropriate response; and (e) Quickly deliver a clear, brief response (Groves and Couper, 1998; and Groves and McGonagle, 2001). A trainee handbook was developed based on information gathered in the focus groups and included a description of that 5-step refusal aversion process, a

1 This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau Publications. This report is released to inform interested parties of research and to encourage discussion.

catalog of the ten main themes of reluctance for the survey, and a place to record new respondent concerns and response strategies. The book also included descriptions of environmental clues, respondent cues and characteristics that may help interviewers diagnose respondent concerns.

Exercises. A number of exercises were designed to familiarize interviewers with the refusal aversion process and systematically increase their ability to use it in a production interview. Initial exercises deliberatively guided participants through the refusal aversion steps. Concluding exercises focused on increasing the speed and accuracy with which interviewers used the process.

2.3 Experimental Design and Procedure

Equal numbers of the interviewers selected from two census regions, Dallas and New York, were assigned to the experimental and to the control group. The control group received no special training while the experimental group received the refusal aversion training. Thirteen New York interviewers attended an eight-hour training session held August 22, 2001 in Manhattan. Seven Dallas interviewers attended a similar session conducted in the Dallas regional office August 29, 2001.

All interviewers had both pre-training and post-training production sample cases in the NHIS. There were 766 pre-training cases and 1780 post-training cases. Analyses are derived from NHIS production cases of these 40 interviewers from July 2, 2001 through the collection period beginning December 24, 2001.

Unlike earlier tests, survey supervisors rather than researchers conducted the refusal aversion training. Prior to the training, researchers held a half-day train-the-trainer seminar that included the history and purpose of the training and a walk-through of the training design.

3. Results

3.1 Interviewer-level Cooperation Rates

Interviewer-level mean cooperation rates for pre-training and post-training periods appear in Table 1. "Cooperation" is defined as the ratio of complete and partial interviews made by the interviewer who first contacted the case to their total number of first contacts. "Contact" means the interviewer made face-to-face contact with an eligible member of the household. A contact was determined using the last observation from the first interviewer's CAPI call history record where the outcome code indicated one of the following had occurred: a completed interview, a partial interview where no follow-up was needed, an insufficient partial interview, a sufficient partial interview where follow-up was needed, an insufficient partial interview, refusal, or "Other Type A" interview as defined in the list of NHIS interview outcome codes.

Table 1. Interviewer-level Mean Cooperation Rates[a], Pre-training and Post-training Period, National Health Interview Survey (unweighted)

		Mean Cooperation Rate				Difference	
		Pre-Training		Post-training			
	No. of Interviewers	Mean %	SE	Mean %	SE	Difference	SE[c]
New York							
Control Group	13	87.2	(2.4)	84.8	(1.7)	-2.4	(3.0)
Experimental Group	13	79.4	(3.6)	86.1	(1.8)	6.7	(4.1)
Net difference						9.1[b]	(5.0)
Dallas							
Control Group	7	91.6	(1.9)	86.3	(1.7)	-5.3	(2.6)
Experimental Group	7	84.8	(2.9)	90.5	(1.8)	5.7	(3.4)
Net difference						11.0[b]	(4.3)
Total							
Control Group	20	88.8	(1.7)	85.4	(1.2)	-3.4	(2.1)
Experimental Group	20	81.3	(2.4)	87.6	(1.3)	6.3	(2.7)
Net difference						9.7[b]	(3.5)

[a] Data for period beginning July 2, 2001 through December 31, 2001, are used to calculate the unweighted mean of interviewer-level cooperation rates as follows: the number of complete and partial interviews by the first interviewer divided by the number of households contacted by the first interviewer.

[b] Differences between pre-training and post-training cooperation rates between treatment groups are significant at the .05 level.

[c] Standard errors reflect clustering of observations into interviewer assignments.

Large within group differences are noted between pre-training and post-training periods. The experimental group's cooperation rate increased 6.3 percentage points while the control group's decreased 3.4 percentage points. Though the difference in the control group for the combined total does not reach significance at the .05 level, it is comparable to the downward trend in cooperation rates among the remaining regions during the 3rd and 4th quarters. The decrease may be due to a seasonal variation in respondents' ability and willingness to participate in a 70 minute interview. NHIS response rates tend to decline in the fourth quarter of the calendar year (Riddick, personal communication).

3.2. Evaluation of Training Absorption

The evaluation of the individual interviewer's retention of the training was based on a brief, rapid-fire type of interaction with a confederate researcher who posed as a reluctant respondent. Twelve utterances were based on the ten concerns outlined in the refusal training handbook described earlier. The confederate respondent posed the following twelve concerns to the interviewer: (1) We already mailed back our census form. (2) So what is this all about anyway? (3) Well, I'm just on my way out right now. (4) I really don't have time for all of this. (5) I just answered a survey last week. (6) I don't think my health is anyone's business but my own. (7) So who else gets this information? (8) I think the government has all the information they need about me already. (9) Why do you have to interview me? (10) The letter says that this is a voluntary survey, so I don't have to do this, right? (11) I don't understand why you want me to sign this consent form. (12) I don't have any health problems; talk to somebody else. Utterances were made by the confederate in the same order and manner for each interviewer. Interviewers were aware that this was a test of their training retention. Evaluations were audio taped.

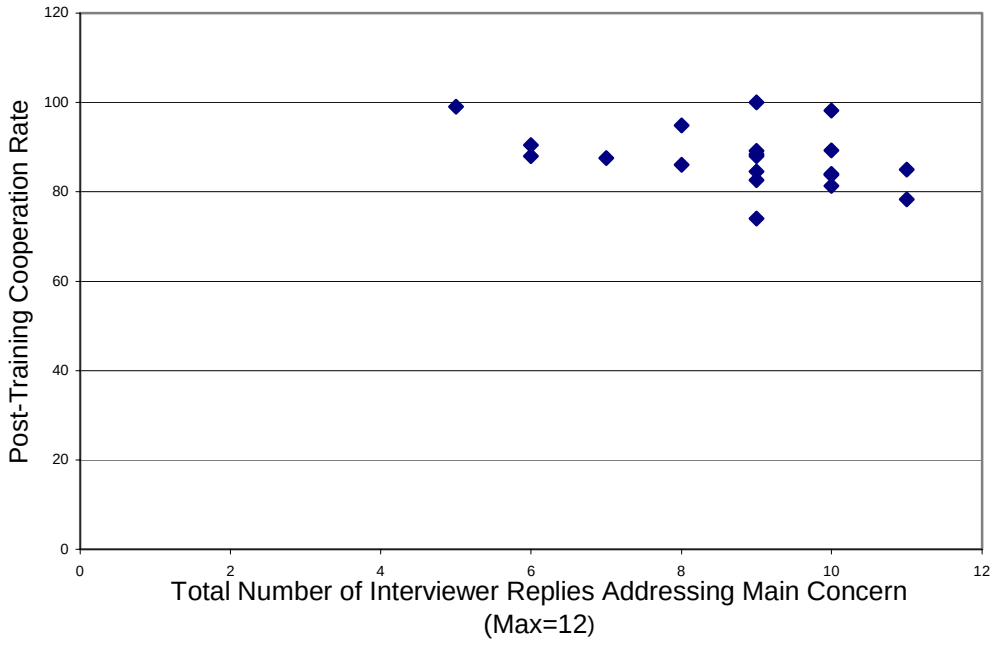
Four raters listened to the evaluation interactions and coded each of the twelve interactions across five main criteria outlined by Groves and Couper (1998); and Groves and McGonagle (2001). The five basic steps to encouraging a survey respondent were conceptualized by five dichotomous measures including: an immediate response by the interviewer to the respondent's concern (NO PAUSE), an interviewer's delivery of the first correct response to the respondent's concern (ADDRESS MAIN CONCERN), an interviewer's delivery of only one main theme regardless of whether or not it is correct (USED ONE THEME), an interviewer's use of softening statements and an empathetic tone (EMPATHY) and an interviewer's use of a confident and professional tone (CONFIDENT). In addition to these five measures, the total length of intermediate pauses, total length of each interaction, and total exercise time were recorded by each coder. Although four coders evaluated the interviewers, this paper uses data from only the first coder. Future analysis will use data from the other three coders to evaluate intercoder reliability.

Table 2 is a correlation matrix of the interviewers' post-training cooperation rates and their coder assigned scores across the 12 concerns. All five indicators show the expected positive correlations but the magnitudes are quite small. The matrix also shows that the five criteria are highly correlated with each other. The five criteria may be affecting the same construct. For example, interviewers who are confident do not pause. From the correlation matrix, "No Pause" and "Address Main Concern" showed the highest correlation with interviewer post-training cooperation rate. Figure 1 is a scatterplot showing the correlation between the index of addressing the main concern and the interviewer's post-training cooperation rate. The failure to construct an evaluation tool that strongly predicts post-training performance continues to plague this effort, as was true in the Groves and McGonagle work (2001).

Table 2. Correlation Matrix of Post-Training Evaluation Average Scores on Five Dimensions
(20 Interviewers)

	Post-Training Cooperation Rate	No Pause	Address Main Concern	Used One Theme	Empathy	Confidence
Post-Training Cooperation Rate	1.00	0.12	0.10	0.05	0.01	0.05
No Pause		1.00	0.97	0.88	0.92	0.90
Address Main Concern			1.00	0.92	0.92	0.89
Used One Theme				1.00	0.82	0.86
Empathy					1.00	0.79
Confidence						1.00

Figure 1. Scatterplot of Post-Training Cooperation Rate by Total Number of Interviewer Replies Addressing Main Concern



3.3 Multivariate Modeling

The use of multivariate modeling, in the context of a random effects model, permits simultaneous estimation of the experimental effects, the marginal impact of the evaluation score, and proper model-based estimates of standard errors. Two different models were estimated for the logit of the cooperation propensity, p_{ij}, the probability of obtaining an interview from a contacted respondent to the survey. The first model attempted to measure the effect of training on the difference between the pre-training cooperation propensity and the post-training cooperation propensity:

$$\ln\left[\frac{p_{ij}}{(1-p_{ij})}\right] = \beta_{0_j} + \beta_{1_j}(POST_{ij}=1) + \varepsilon_{ij}$$

$$\beta_{0_j} = \gamma_{00} + u_{0j}$$

$$\beta_{1_j} = \gamma_{10} + \gamma_{11}(TRAIN_j = 1) + u_{1j}$$

where $POST_{ij}$ equals 1 for the *i*-th contacted respondent assigned to the *j*-th interviewer if *i* is part of the sample released after the training (0, otherwise), and $TRAIN_j$ equals 1 if the *j*-th interviewer was assigned to the training group (0, otherwise).

The second model attempts to separate the base effects of training from those effects associated with measured differential absorption of the training guidelines, as measured by a computed summary evaluation score (0-60) based on coding the audio-taped evaluation session, $TEST_j$, measured on each of the interviewers.

$$\beta'_{1_j} = \gamma'_{10} + \gamma'_{11}(TRAIN_j = 1) +$$

$$\gamma'_{12}[TEST_j \mid (TRAIN_j = 1)] + u'_{1j}$$

To the extent that γ'_{12} is a large positive number, those interviewers who scored well on the evaluation experience greater gains in cooperation propensity.

Models were fit separately for the two regional offices and for the pooled sample. Estimated model parameters appear in Table 4. The "base model" measures the direct effect of the training. In both regional offices the direction of the training effect is to increase cooperation (coefficients of .67 for Dallas and .05 for New York), but based on the number of interviewers used in the experiments, the effect sizes do not reach traditional levels of statistical significance. When the data are pooled across the two sites, the training effect is positive (.34) and achieves a p-value of .14. This is largely consistent with the discrete data analysis of Table 1, but we prefer the variance estimation reflecting the variation in workload size among interviewers.

There are other coefficients in the base model that are noteworthy. The coefficient labeled, "Post-Pre Difference, Intercept, (γ_{10})," can be interpreted as the expected

difference for cases assigned to the control group interviewers. Note that the coefficient is negative (but with a high standard error) for Dallas and positive for New York (also with a large standard error). The differences in Table 1 for the control group interviewers were both negative, implying declines in cooperation between the dates before and after training. The signs of the coefficient for New York reflect the weighting by interviewer workload in the model-based estimates.

Moving to the Evaluation Score Model, we examine first the coefficient for the score on the evaluation measure. Interviewers were scored across five criteria for twelve concerns. The average score among the trained interviewers was 43.1. We hypothesized positive coefficient values, reflecting the tendency for those interviewers scoring high on the evaluation to achieve higher gains (post-pre differences). This is unsupported by the data, with large negative effects among the Dallas interviewers (-.13, standard error = .04) and negative, but negligibly so in New York (-.02, standard error = .02). The pooled estimate of the evaluation score effect is -.03, with a standard error of .02.

The evaluation score is intended to represent the knowledge relevant to the training protocol exhibited by the interviewer immediately after training. This is clearly a proxy for knowledge *gained* in the training. To the extent that interviewers who scored high on the evaluation already possessed the revealed knowledge prior to training, we would expect the post-score coefficient to be biased. This is our current interpretation of the finding of negative effects of the evaluation score.

In summary of the multivariate modeling, the training effects appear to be positive but too small to be reliably detected with the number of interviewers studied. The evaluation protocol is not itself a useful predictor of the *gains* in cooperation due to training, but the interpretation of that finding is complicated by no pre-training skill measurement.

Table 4. Random Effects Logistic Regression Coefficients for Cooperation Propensity, Base Model and Evaluation Score Model, Separate Regional Office Models and Pooled Model

	Base Model			Evaluation Score Model		
	Coeff	SE	p	Coeff	SE	p
New York Regional Office, Fixed Effects						
Base Propensity, Intercept, (γ_{00})	1.75	0.22	0.00	1.78	0.22	0.00
Post-Pre Difference, Intercept, (γ_{10})	0.11	0.24	0.66	0.11	0.25	0.66
Training (TRAIN=1), (γ_{11})	0.05	0.30	0.86	0.93	1.06	0.39
Evaluation (TEST|TRAIN=1), (γ'_{12})				-0.02	0.02	0.40
Variance component (Var(u_1))	0.048			0.096		
Dallas Regional Office, Fixed Effects						
Base Propensity, Intercept, (γ_{00})	2.23	0.30	0.00	2.22	0.29	0.00
Post-Pre Difference, Intercept, (γ_{10})	-0.44	0.37	0.19	-0.40	0.29	0.20
Training (TRAIN=1), (γ_{11})	0.67	0.34	0.07	6.51	1.90	0.006
Evaluation (TEST|TRAIN=1), (γ'_{12})				-0.03	0.02	0.17
Variance component (Var(u_1))	0.483			0.453		
Pooled Sample Fixed Effects						
Base Propensity, Intercept, (γ_{00})	1.97	0.19	0.00	1.96	0.19	0.00
Post-Pre Difference, Intercept, (γ_{10})	-0.16	0.20	0.43	-0.15	0.20	0.45
Training (TRAIN=1), (γ_{11})	0.34	0.22	0.14	1.59	0.92	0.10
Evaluation (TEST|TRAIN=1), (γ'_{12})				-0.03	0.02	0.17
Variance component (Var(u_1))	0.309			0.328		

4. Summary and Discussion

The purpose of this study was to improve the theoretically based refusal aversion training by implementing it in a face-to-face demographic survey. Groves and Couper (1998) derive their theory of householder participation from this context and embody it in the aversion process described in the training.

The evaluation tool used here was reformulated from its original design in the Groves and McGonagle work (2001). Without a pre-training measure of skill, however, we cannot identify a change in performance level from the pre-training to post-training period in line with the change in cooperation rates. Future work might include such a measure and use field staff as evaluators. With four coders having participated in the scoring, we can now investigate the intercoder reliability of the evaluation measures.

In developing the refusal aversion materials for the NHIS, we identified a few themes of concerns common to other survey requests. Interviewers, however, continue to provide responses and strategies that are unique to specific surveys based on the study population, the nature and timing of the request, the length or mode of data collection, and so on. This affirms our understanding that some portions of the classroom materials must be redeveloped for individual surveys. In addition, although focus groups provide a multitude of verbatim concerns and response strategies, omissions or changes in the survey-taking climate might weaken the validity of the materials over time. Finally, the Census Bureau could frontload information onto laptop interview cases such as neighborhood characteristics or information from previous contacts to the household. This would help interviewers further prepare their approach for particular households.

These results mirror findings in previous research (Shuttles, Welch, Hoover, and Lavrakas, 2002; Groves and McGonagle 2001; Mayer and O'Brien 2001). The theoretical construct driving this refusal aversion training protocol has proved valid across survey organizations, survey topics, populations and modes. Interviewers feel their work requires skills taught in this training, feel they need more of such training to succeed (Doughty, Foley, Spuches and Yonai 2001), and feel this training fills that gap.

When cooperation rates between pre-training and post-training phases are compared in this study, the positive effects of training persist although they do not meet traditional levels of statistical significance in the multivariate models. Establishing statistical significance with a small number of interviewers proves difficult when accounting for varying workload sizes, as shown between Table 1 and the multivariate analyses. Workloads vary for a number of reasons but are in part due to the NHIS sample design that over-samples special populations. Piloting this training in several national surveys would permit a fuller evaluation at the scale in which it will ultimately be used, provide the best mechanism for a statistical understanding of its effects, and yield operational information managers should find useful in choosing among strategies for managing unit nonresponse.

References

Couper, M. P., & Groves, R. M. (1992). The role of the interviewer in survey participation. Survey Methodology, 18, 263-277.

Doughty, P., Foley, A., Spuches, C., & Yonai, B. (2000). Current Population Survey interviewer training evaluation study: Final Report. Report submitted to the US Bureau of the Census, January 2000.

Encouraging Response in QDERS: Persuading the Reluctant Respondent: Reference Guide. Prepared by the Center for Survey Methods Research at the United States Census Bureau, Spring/Summer 2000.

Encouraging Response in the NHIS: FR Handbook. Prepared by the Center for Survey Methods Research at the United States Census Bureau, August 2001.

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. Public Opinion Quarterly, 56, 475-495.

Groves, R. M., & Couper, M. P. (1998). Nonresponse in household interview surveys. New York: John Wiley.

Groves, R. M., & McGonagle, K. (2001). A theory-guided interviewer training protocol regarding survey participation. Journal of Official Statistics, 17, 249-265.

Mayer, T. S., & and O'Brien, E. M. (2001). "Interviewer Refusal Aversion Training to Increase Survey Participation," 2001 Proceedings of the American Statistical Association, Survey Research Methods Section [CD-Rom], Alexandria, VA: American Statistical Association.

Oksenberg, L., Coleman, L., & Cannell, C. F. (1986). Interviewers' voices and refusal rates in telephone surveys. Public Opinion Quarterly, 50, 97-111.

Riddick, H. (personal communication, February 22, 2002)

Shuttles, C. D., Welch, J. S., Hoover, J. B., & Lavrakas, P. J. (2002, May). "The development and experimental testing of an innovative approach to training telephone interviewers to avoid refusals." Presented at the 57th Annual American Association for Public Opinion Research Conference, St. Petersburg, FL.

Acknowledgements

We express our gratitude to various reviewers and collaborators, especially Nancy Bates, Angel Broadnax, Gerry Burt, Joseph Cavey, Allison Cenac, Lester Farthing, Toni Lubich, Jan McStay, Al Mirabel, Adrienne Oneto, Blair Russell, Lee Siciliano, Linda Schlueter, and Tommy Wright.