An Inquiry into the Efficacy of a Complementary Training Session for Telephone Survey Interviewers

Claire Durand, Marie-Eve Gagnon, Christine Doucet and Eric Lacourse

Electronic version

URL: <http://journals.openedition.org/bms/537>

ISSN: 2070-2779

Publisher

Association internationale de méthodologie sociologique

Printed version

Date of publication: 1 October 2006

Number of pages: 5-27

ISSN: 0759-1063

Electronic reference

Claire Durand, Marie-Eve Gagnon, Christine Doucet and Eric Lacourse, « An Inquiry into the Efficacy of a Complementary Training Session for Telephone Survey Interviewers », *Bulletin de méthodologie sociologique* [Online], 92 | 2006, Online since 01 October 2009, connection on 21 April 2019. URL : <http://journals.openedition.org/bms/537>

This text was automatically generated on 21 April 2019.

© BMS

An Inquiry into the Efficacy of a Complementary Training Session for Telephone Survey Interviewers

Claire Durand, Marie-Eve Gagnon, Christine Doucet and Eric Lacourse

Introduction

- 1 Whatever the setting in which it takes place, the training of employees must take a number of contingencies into account. Organizations see training as an investment, and after making such an investment, they hope to retain their employees; therefore, they usually train permanent employees with the stated goal of improving productivity.
- 2 The training of interviewers working in survey research often takes place in an employment context with few training incentives. Generally, there is a high level of turnover, and many employees work part-time or on call. Interviewers may work for more than one employer and go from project to project. To a certain extent, they may be able to negotiate how many hours they work and when. In a sense, their situation is similar to that of contractual employees.
- 3 In the context of social research, interviewers are asked to carry out an activity that is considered to be scientific in nature. However, in private firms, there is a trade-off to be made between quality and cost-effective production; employees must be competent and flexible but not overly expensive, keeping in mind the costs associated with recruitment and training. Against this background, the research team devised a short training session that focused on improving cooperation rates and that complemented the initial training provided by a private survey firm specialized in social research.

The Context: The Training of Interviewers

- 4 A review of the literature specific to interviewer training turns up a surprisingly low number of publications dealing with the question. A number of authors (Mayer and O'Brien, 2001; O'Brien *et al.*, 2002; McConaghy and Carey, 2004; Cantor *et al.*, 2004) have experimented with training along the lines presented by Groves and McGonagle (2001). This type of training, called ART (avoiding refusal training), focuses on tailoring and maintaining interaction. Briefly, the various phases consist in gathering householders' concerns from expert interviewers, classifying these concerns by theme, examining with experienced interviewers how to respond to the different types of concerns, and writing a "manual" that serves as the basis for training. The training itself lasts about eight hours. Interviewers are taught to identify concerns, classify them appropriately, learn the best reaction to each type of concern and, finally, actively practice how to react. There are variations among researchers, but the basic pattern remains the same. Some researchers incorporate measurement instruments (pre- and post-training evaluation, personality tests, observation scores, etc.) to better evaluate what was learned during training and relate it to subsequent performance.
- 5 The two experiments presented by Groves and McGonagle (2001), conducted by telephone, were follow-ups, one among contact persons in businesses, the other among farm operators. Interviewers were experienced, permanent interviewers working for governmental organizations (U.S. Bureau of Labor statistics, National Agricultural Statistics Services). Groves and McGonagle reported a 14-percent improvement in cooperation following training; however, this increase was not clearly related to performance in training. They also concluded that they were "skeptical of a naive application of these findings to RDD household telephone surveys because of the tendency for radically reduced length of interactions prior to a householder's decision."
- 6 However, Mayer and O'Brien (2001) did exactly that; *i.e.*, they followed up on Groves and McGonagle (2001) in the context of an omnibus RDD telephone survey of the general population. They had 24 interviewers, with between five months and five years of call-centre experience, distributed among three groups: a control group, which received no training; a "before" group, which received the eight-hour training session before the beginning of the project (two two-week surveys); and a "between" group, who received the training session between the two surveys. They reported an improvement of between 3 and 7 percent in first contact cooperation rate for interviewers who participated in the training and up to 14 percent over time compared with those who did not receive the training. They recommended that follow-up research use a) a larger sample of interviewers, b) a longer data collection period, c) baseline data for participant groups, and d) the design of an interviewer evaluation as a management tool. Their subsequent experiment (O'Brien *et al.*, 2002) was carried out in a face-to-face context (National Health Interview Survey). They concluded that the training did have an effect, but they did not find significant relationships between post-training evaluation scores and first contact cooperation rates.
- 7 McConaghy and Carey (2004, 2005) also followed up on Groves and McGonagle (2001) in an experiment with face-to-face interviews conducted for the General Household Survey in the UK's Office of National Statistics. Low performers were targeted for either an ART training or a placebo-like "Day Away" intervention. They obtained results similar to

Groves and McGonagle (2001): some improvement in response rates (around 9 percent). However, the ART trainees could be divided into improvers (n=9) and “no change” (n=6). The improvers experienced a 15-percent improvement in response rate following training; response rate tended to decrease slightly over the long term but remained higher than before training. Those who improved most during training tended to achieve a better response rate after training. Comparing the ART training with the placebo “Day Away” group confirmed that ART training was more efficient than a mere placebo intervention. Finally, Cantor *et al.* (2004) converted this training program into a computerized format. They conducted their experiment in the context of a follow-up of non-respondents to a mail survey. Although they found differences between the control group and the two experimental groups, none of these differences were significant.

- 8 In summary, authors using the ART program usually found some improvement in performance among interviewers following the training, but these differences were not always statistically significant and could rarely be attributed to knowledge acquired during training. There is also the possibility of a Hawthorne effect; *i.e.*, that the mere attention paid to interviewers explains the subsequent improvement in performance. However, McGonaghy and Carey’s (2005) design allows to rule out this possibility in their case. Only one of the experiments was conducted in the context of a RDD telephone survey (Mayer and O’Brien, 2001), and in most experiments, the pool of interviewers was rather homogenous, experienced, and small. Finally, in one case (Groves and McGonagle, 2001), interviewers did not seem to enjoy the training while in another case (McGonaghy and Carey, 2005), they did.
- 9 Other attempts at improving training have been carried out mostly in Belgium and in the Netherlands in varying contexts. Carton (2000) studied training in the context of five different projects carried out by ISPO (KU Leuven). She concluded that a) there was in fact very little difference in performance between interviewers working on the same project, b) interviewer performance may vary between projects, and c) training should be done differently and separately for experienced and newly hired interviewers, building on acquired experience with the former, while starting with theoretical explanations before going on to practical exercises with the latter. She also advocated that in the selection process, interviewers be better informed of what is expected of them from a “cognitive, communicational and social point of view.”
- 10 Dijkstra and Smit (2002) recorded interactions in an attempt to understand what works and what does not. The interviewers were all female students in a department of communication with no interviewing experience, making the generalizability of the data doubtful. However, it is interesting to note that a) they saw no difference in efficiency between tailoring and persuasion attempts, and b) maintaining interaction was considered effective only when sustained by the respondent, rather than the interviewer. Finally, Van der Zouwen, Dijkstra and Smit (1991) reported that when interviewers are trained to adopt a more formal style, they tend with time to return to a more personal style.
- 11 Summarizing these various experimental results, one could conclude that the impact of training appears to be generally low. This could be due to the content of the training and the methods used. Or it could be due to an inappropriate setting such as insufficient variance in the performance measured due to type of interviewers (experienced and generally high performers), type of project (easy follow-ups as opposed to RDD surveys of the general population) or project length (too short to enable selection of interviewers,

training and assessment of impact). In addition, the measure of performance generally used -- cooperation rate at first contact -- may be unreliable, thus preventing an effect from being identified (Durand, 2005). Finally, it is rather surprising that no longitudinal analysis of the evolution of performance has been performed in such contexts. Before concluding that performance in this field is a question of predispositions that interviewers either have or lack and which cannot be improved by training, it appeared necessary to tackle these four possible reasons for the low impact of training; *i.e.*, setting, content and methods used, measurement of performance, and analysis of performance.

- 12 The goal of this study was to devise a short complementary training of interviewers in the context of a social research and of a RDD-based telephone survey. At the same time, we wanted to be as close as possible to the situation faced by private firms; *i.e.*, a substantial proportion of new interviewers, a high turnover, a financial context that prevents devoting much money to training. Inspired by literature on empowerment and self-efficacy, and taking into account the means at our disposition, we decided to devise a cognitive training focused on knowledge acquisition. This training should be viewed as a first experiment that ought to be replicated and improved.

The Experiment

Setting

- 13 The experiment took place in the context of the Canadian Addiction Survey of 2003–2004. The survey used a stratified sample of the provinces and territories of Canada. It was conducted from a single site located in Montreal, Canada,¹ by a private survey company. The fieldwork began on December 16, 2003, and ran for one week; after the Christmas break, it started up again on January 9 and finished on April 19, 2004. Close to 90 percent of the interviews were conducted in English, the rest in French. The response rate (AAPOR RR3) was 47 percent for a total of 13,999 completed interviews. The average interview length was 24.6 minutes.
- 14 Overall, 79 interviewers worked on the project for a total of 2,436 interviewer-days; interviewers worked a maximum of 65 days, and the survey was in the field for 109 days. Since the firm usually conducts most of its interviews in French, it had to recruit a substantial number of new, bilingual or English-speaking interviewers for the project.¹ A questionnaire pertaining to interviewers' attitudes, behaviours and characteristics (Lemay and Durand, 2002) was distributed to interviewers as soon as they had worked at least 20 hours on the project. In all, 72 interviewers worked at least 7 days on the project; 57 of them completed the questionnaire, and 54 gave their ID codes, allowing us to match their questionnaire with performance data.
- 15 Table 4 paints a portrait of the interviewers who answered the questionnaire. One-third of the interviewers were younger than 30 years old, while 39 percent were 40 years or older, with the mean age being 37 years old. There was an almost even split between men and women, and 75 percent had French as their only mother tongue. Interviewers had nearly three years of experience on average as interviewers. However, 72 percent (n=41) were newly hired by the firm, and 14 of them were new to the job of interviewing. Most interviewers (84 percent) considered themselves regular employees, split almost equally between full-time and part-time. More than one-third (36 percent) worked 25 hours or less per week, while one in four (23 percent) worked more than 35 hours a week.

Content and Methods

- 16 The literature on training in organizations and in the survey field, as well as observations within the firm, directed us toward a cognitive type of training that would complement the basic training given by the firm which was more practical and concrete. The assumption was that knowledge would translate into the development of abilities because interviewers would better understand why they were asked to do within-household selection, convince selected respondents, and perform refusal conversion, and what happened when they tried to carry out these tasks. The goal was to increase interviewers' self-confidence and intrinsic motivation in order to help them find their own solutions to the problems they faced; *i.e.*, to tailor their methods according to their own personal style of interaction. We postulated that tailoring was related not only to respondents but also to interviewers' habits in dealing with people. If interviewers' reactions to respondents seemed artificial or pre-programmed, it may not work.
- 17 Two types of information were identified as lacking and possibly helpful to interviewers. First, concrete information on sampling in general, and on the particular sampling frame used for surveys of the general population, was meant to improve interviewers' understanding of why they have to select respondents and convince those selected to answer the survey. The second type of information, related to why people refuse to participate in surveys and how experienced interviewers deal with refusals, should allow for a better understanding of the social interaction and of what might be happening in the household they are calling. The stated goal was to reduce the stress associated with "taking it personally," *i.e.*, taking too much responsibility for the refusals and therefore becoming paralyzed by stress. We rejected practice because of the time needed to use that type of training in an efficient way and because of the "factice" nature of practice in the context of telephone surveys where interviewers have to improvise and react very quickly.
- 18 One consideration in setting up the training session was deciding who to train. Since training is likely to be more efficient and relevant with low performers (Groves and McGonagle, 2001) and since the focus of training should differ according to performance and experience (Carton, 2000), it was decided to focus on low performers and newly hired interviewers.
- 19 The second decision concerned a feasible setup for the training. We decided to offer short training sessions because in the context of a private firm's activities, this seemed more realistic to implement in various situations. In addition, a short session has the advantage of making it easier to keep interviewers concentrated and interested. It was thus decided to hold three one-hour sessions during the same afternoon and to target the low performers and newly hired interviewers working on that particular day. The other low-performers and newly hired interviewers would act as a control group. This way of selecting appeared the most feasible way to proceed since the characteristics of interviewers are not normally related to which day they work. When the training sessions were scheduled, one-third of the fieldwork remained to be done, and the type of work left was becoming more specific (mostly follow-ups on appointments and refusal conversion).

The Training

- 20 The first step in the process was to identify the low performers. During the first weeks of the fieldwork, the administrative database² was collected every two weeks to monitor the evolution of performance. Performance itself was measured using an index of the net contribution to performance (NCPI³: Durand, 2005). This index takes into account all the tasks performed by interviewers, including follow-ups on appointments and refusal conversion; it can be computed whatever the task performed, contrary to cooperation rate at first contact (Durand, 2005). Semi-parametric group-based trajectory analysis (Nagin, 1999) was used. This method allows for the identification of different trajectories of the evolution of performance. Two groups, which may be termed high and low performers, were identified. Figure 1 illustrates the evolution of performance from day 7 in the field (after the Christmas break) up until two weeks before training, when interviewers were selected. The analysis identified 25 interviewers in the low-performance trajectory and 42 in the high-performance trajectory. The trajectories of low versus high performers differed in the average starting level (-.87 compared with -.40). The average rate of improvement was similar (.0159 per day in the field) but the average daily performance of the high performers appeared to become more stable over time.

Figure 1: Trajectories of low and high performers two weeks before training

- 21 The field director was asked to similarly classify the interviewers into two groups. There is an inter-rater agreement of .58 between the two classifications. The classification of performance trajectories was used to decide who to train.⁴ In addition, 10 interviewers, hired between the identification phase and the training session, were added to the group to be trained. Table 4 shows that there is little difference in the characteristics of respondents to the questionnaire given to interviewers when comparing the newly hired,

the low performers and the high performers or the trained and untrained interviewers. The newly hired tended to be younger and were more likely to be women.

- 22 All the identified low performers and newly recruited interviewers (n=19) working on the day chosen for training were invited to participate in one of the training sessions. These sessions were scheduled to accommodate the different work schedules of interviewers. All the selected interviewers accepted the invitation and were paid during the training. One interviewer was absent and one was late and transferred to the next session. Therefore, the three sessions were given to a group of six (five low-performers and one newly-hired), a group of five (two low and three new) and one of seven interviewers (six low and one new). The first one-hour session was followed by a one-hour break, during which the two assistants and the principal researcher/trainer had a debriefing. This discussion allowed for an adjustment of the two following training sessions, which were held back-to-back.
- 23 The first part of the training pertained to sampling. To liven up the training, bags of M&Ms were used (Auster, 2000) to explain different notions in a concrete manner: a) What is a sample? -- each M&M bag as a sample, b) What is the effect of non-random non-response on the sample? -- example with refusals and with selection within household.
- 24 The second part of the training was less lively. No screen or LCD projector were available. Therefore, reproductions of transparencies were distributed. These transparencies presented results from research on refusals (mostly from Goyder (1988), *The Silent Minority*) and information gleaned from interviews of high performers on how to deal with refusals. This part essentially comprised presentation of information and discussion.
- 25 The trainees seemed to appreciate the first part and asked questions such as: What is a sample, really? How is the sample produced? How are the telephone numbers selected? Why can't I interview any person who is willing to answer? Why do I have to interview a person who is not interested? Why ask an elderly woman if she has ever smoked marijuana? We answered these questions, sticking to the M&M analogy and the percentage of yellow candies in the bags. Since the interviewers were working on the Canadian Addiction Survey, the analogy was made relating the yellows to marijuana smokers in order to explain that the proportion of marijuana smokers in the population would not be correct if the non-smokers were not interviewed. As for the second part of the presentation, some interviewers seemed very surprised by the information, which went opposite to what they believed. Finally, the next day, the field director told us that everybody was talking about the M&M experiment and that it was a "critical success", which is, at least pedagogically, a good sign.

Results: Impact of Training

- 26 Training can improve performance, which is the first goal usually pursued by organizations who use it. It may also have other, less spectacular but nonetheless substantial positive effects. For instance, it may help retain personnel who would otherwise have left. It may also improve self-confidence and reduce job-related stress, therefore helping to maintain a good work climate and eventually improving the performance of interviewers as a whole. These potential impacts are assessed below.

a) Knowledge Acquisition and Stress Reduction

- 27 In order to assess the impact of the training, a questionnaire was distributed to all the interviewers still with the firm two weeks after the training. Version A, for interviewers who were not invited to a training session, comprised eight questions pertaining to beliefs regarding reasons for refusals, perception of control, self-attribution for performance, and perception of their understanding of sampling and within-household selection. Version B, distributed only to the trained interviewers, had an additional section concerning the evaluation of the training *per se*; i.e., whether the training helped them understand sampling and selection and find ways to convince respondents. In all, there were 18 interviewers in the trained group, but one did not return after the training (13 respondents); 21 in the control group, but only 7 still working on the project on training day (4 respondents); and 42 in the high performers group, with 30 still working on the project on training day (26 respondents).
- 28 Table 1 presents the evaluation of the training. Items can be divided into three blocks. The highest agreement block: Respectively 12 and 10 interviewers (of 13) fully or somewhat agreed that the training helped them understand sampling and within-household selection. This refers to the part where we used the M&M experiment. The second-highest agreement block: Respectively nine of 13 and eight of 12 respondents fully or somewhat agreed that the training helped them understand the reasons for refusals and why it is important to convince potential respondents to cooperate. The low agreement block: Respectively seven and six interviewers of 12 respondents fully or somewhat agreed that the training helped them find arguments to convince respondents or feel more comfortable with convincing. This part refers to the transfer of acquired knowledge into practice. This aspect was not part of the training *per se*.

Table 1: Evaluation of the Training Session

	Total N	Fully agree N	Somewhat agree N	Fully or somewhat disagree N
Training helped understand how phone numbers are selected	13	8	4	1
Training helped understand selection within households	13	6	4	3
Training helped understand the reasons for refusals	13	4	5	4
Training helped understand why it is important to convince people to cooperate	12	3	5	4
Training helped find arguments to convince respondents	12	2	5	5
Training helped feeling at ease with the task of convincing respondents	12	1	5	6

- 29 Table 2 shows the distribution of answers to questions pertaining to the interviewer's role and to knowledge regarding sampling and within-household selection. The distribution of answers is highly skewed in favor of the two most favorable categories. Table 2 shows that the levels of high agreement for the three first interviewer-related items varied between 48 percent (20 of 42 felt they controlled the situation during interviews) and 60 percent (26 of 43 fully agreed that people accept because the interviewer is self-confident). In addition, 33 percent (14 of 43) said they were capable of finding arguments most of the time, and 60 percent that the role of the interviewer in the decision to participate is very important. A scale of these items may be computed.⁵ When the trained interviewers were compared to the high performers group -- interviewers from the control group being too few to make a significant comparison -- on this scale, the high performers appeared more likely to be self-confident and attribute cooperation to their own work ($F=6.23$, $p=.017$) than the trained interviewers.

Table 2: Interviewers' Self-perception of Their Role, Behaviour and Knowledge

		Total	Very much agree	Somewhat agree	Somewhat disagree
		N	N	N	N
Interviewer-related items	Feeling of control over interview	42	23	18	1
	People accept because interviewer makes good impression	42	20	22	0
	People accept because interviewer is self-confident	43	26	16	1
			Most of the time	Often	Sometimes or rarely
	Able to find convincing arguments	43	14	22	7
			Very Important	Somewhat important	Not much important
	Influence of interviewer in decision to participate is...	43	26	15	2
	Scale Mean (scale of 1-3, 1=high)	12 25	Trained: 1.8 (s.d. .39) High-performers: 1.4 (s.d. .40)		

			Very well	Rather well	Not well
Knowledge of sampling and selection	Know why selection within household is done	43	35	8	0
	Know how telephone numbers are selected	43	21	13	9
	Scale Mean (scale of 1-3, 1=high)	13 26	Trained: 1.4 (s.d. .46) High-performers: 1.5 (s.d. .53)		

			Has improved much	Has improved somewhat	Remained stable
Confidence in capacity to convince	Trained	13	7	6	0
	High performers	26	11	8	7

- 30 Forty-nine percent (21 of 43) said they knew very well how telephone numbers are selected and 81 percent (35 of 43) knew very well why within-household selection is done. Analysis showed that high performers and trained interviewers did not differ in their self-reported knowledge⁶ of sampling and selection ($F=.446$, $p=.51$).
- 31 Finally, trained interviewers were more likely than high performers to declare that their self-confidence had improved since the beginning of their work on the project (Lr test= 6.47 , $p=.039$). All the trained interviewers said their capacity to convince had improved much (7 of 13) or somewhat (6), while 19 of 26 high performers said their performance had improved. The rest (7) stated that it had remained stable.

b) Employee Retention

- 32 Training may also have an impact on employee retention. Low performers and newly hired interviewers usually find the work they do rather difficult and stressful at first. Training may encourage them to persevere: it sends the message that the employer cares about them and wants to retain them, and it stresses -- in the case of our experiment at any rate -- the importance of interviewers in the survey process, while at the same time trying to convey that interviewers should not take all the blame for refusals.
- 33 Figure 2 shows the distribution of the number of days the interviewers still with the firm on training day stayed afterwards. Only one of the 18 trained interviewers left the firm immediately after training. The 17 others worked an average of 18 days (17 days on average if the interviewer who left is included) after the training, with a minimum of four days and a maximum of 34. Of the control group, only seven of 17 were still working on the training date: ten had left between the identification phase and the training day. Among those who could have been trained and were not, the average number of days worked after the training was 11.4, with a minimum of two days and a maximum of 20. In comparison, the 30 high performers who were still working on the project worked an average of 17 days with a minimum of one day and a maximum of 29. However, these differences are not statistically significant ($F=1.35$, $p=.27$).

Figure 2: Employee retention according to training group

c) Performance

- 34 One key element of interest is whether the training had any impact on interviewer performance, restricted in this case to the ability to convince. In order to assess this impact, it was necessary to have a reliable measure of performance and to examine the evolution of performance throughout the project, in order to estimate how the performance evolved after the training session. Performance may be seen as nested in interviewers so that multilevel growth models may be used.

Table 3: Results of Longitudinal Multilevel Analysis of Evolution in Performance

	Model 0 Time linear	Model 1 + time quadratic	Model 2 + group effect	Model 3 + post training time linear	Model 4 + post training time quadratic	Final model without non- sig. Random effects
Fixed effects						
Intercept	-.42092***	-.53231***	-.80286***	-.80620***	-.77552***	-.77178***
- group effect			.45331***	.44612***	.44672***	.44519***
Time linear	.0099***	.01449***	.01525***	.01604***	.01473***	.01462***
Time quadratic		-.00004*	-.000045**	-.00006***	-.00005***	-.00005**
Post training						
time linear				.00789*	.02607***	.02531***
time quadratic					-.0005***	-.0005***
Random effects						
Intercept	.09368***	.09818***	.01996***	.01156**	.01302*	.01624***
time linear	.00001***	.00001***	.00001***	.00001**	.00001	.00001***
time quadratic						
post train. Time linear				.00005*	.00019*	
post train. Time quad.					.00000	
Within interviewer e	.10287	.10194	.10188	.10159	.09972	.10043
Deviance (diff)	1541	1528 (13)	1452 (76)	1433 (19)	1396 (37)	1408(12)
df (diff)	6	7 (1)	8 (1)	12 (4)	17 (5)	10 (7)

- 35 A classical approach to analyzing this type of data is multilevel longitudinal analysis (Singer and Willett, 2003; Hox, 2002; Raudenbush and Bryk, 2002; Snijders and Boskers, 2000) in which the two levels are the interviewer and the daily performance. The analysis did not use the data for each interviewer's two first days of work, since performance is not reliable enough during the initial learning period. As with trajectory analysis, time itself (day in the field) was used as the time variable instead of interviewer experience on the project. Table 4 shows the results for the analyses used to identify a parsimonious and theoretically sound model. The following effects are tested using nested models. First, an unconditional growth model is fit to the data in order to estimate the overall effect of time: a linear effect of time is entered and, since performance can only improve to a certain optimal level, a quadratic component in the effect of time is added. Then, a conditional growth model is tested in order to a) control for the membership in the control group and b) estimate whether training modified the evolution of performance in the trained group. . The final model is parsimonious and keeps only the significant random effects.⁷

Table 4

		Total	Performance group			Training	
			New	Low	High	Untrained	Trained
Age group	18-29	33%	57%	20%	33%	37%	21%
	30-39	29%	29%	40%	24%	24%	43%
	40-49	15%	0%	13%	18%	17%	7%
	50 and over	24%	14%	27%	24%	22%	29%
	N	100%	7	15	33	41	14
Gender	Female	45%	63%	40%	42%	50%	29%
	Male	55%	38%	60%	58%	50%	71%
	N	100%	8	15	33	42	14
University degree	No	54%	50%	57%	53%	56%	46%
	Yes	46%	50%	43%	47%	44%	54%
	N	100%	8	14	32	41	13
Mother tongue	French						
	Other	75%	75%	80%	74%	74%	79%
	N	25%	26%	20%	27%	26%	21%
Status	Regular	84%	100%	92%	76%	79%	100%
	On call	16%	0%	8%	24%	21%	0%
	N	100%	8	13	34	42	13
	Full time	44%	50%	54%	38%	38%	62%
	Part time	56%	50%	46%	62%	62%	38%
Experience as interviewer	0-3 months	25%	13%	20%	20%	26%	21%
	4-12 months	14%	13%	7%	18%	16%	7%
	Over a year	61%	75%	73%	53%	58%	71%
	N	100%	8	15	34	43	14
Perception of job	Complementary	38%	57%	33%	35%	43%	21%
	Transitory	29%	14%	20%	35%	31%	21%
	Permanent	34%	29%	47%	29%	26%	57%
	N	100%	7	15	34	42	14

- 36 Model 0 includes only the linear effect of time. It shows that there was as much variance in performance between interviewers (random variance of the intercept = .09368) as within interviewers between days (.10287). It also shows that the base level of performance improves on average by .0099 per day and that this effect varies between interviewers (since the random effect of time (.00001) is significant). Model 1 adds the quadratic effect of time. This effect is also significant but does not vary between interviewers. Since the variance within interviewers increases to .09818 with this model, it is used as the base model against which the following models are compared. Therefore, the unconditional growth model includes both a linear and a quadratic effect of time.
- 37 In Model 2, membership in the trajectories used for selection is entered in order to control for the pre-training initial status. The between-interviewer variance decreases from .09818 to .01996, meaning that the categorization accounts for 79.7 percent of the variance between interviewers. The intercept (*i.e.*, the starting performance) of the high performers is .446 higher, on average, than that of the low performers.⁸
- 38 The following models test the effect of training. This effect can be modeled in different ways (see recommendations by Singer and Willett, 2003, chap. 6). Training may have an effect on the intercept – a jump in performance following training – or on the post-training evolution of performance. Multiple tests lead us to the final model presented in this paper. In Model 3, the linear effect of time after training is entered.⁹ Model 3 shows that the net improvement of performance after training is .00789 per day ($p=.011$). However, model 4 tells a slightly different story since the quadratic effect of time after training is significant and the linear effect of time is adjusted accordingly. The new model shows a linear effect of .02607 per day after training, to which is added a quadratic negative effect of -.0005, both highly significant at $p=.000$. This means that there is a

linear improvement in performance after training but that the performance eventually reaches a plateau and decreases at the end of the fieldwork. This plateau may be due to the type of work remaining. Final tests of the significance of random effects lead to the final model, in which only the intercept and the overall linear effect of time have significant random effects,¹⁰ which means that there is still some unexplained variation left in the average performance of interviewers and in the average effect of time.

- 39 Figure 3 illustrates the change in the estimated average performance of the three groups; *i.e.*, the high performers, the untrained low performers and newly hired interviewers (control group), and the trained interviewers over the field period, without the first and last weeks, those not being typical. It illustrates the improvement in performance that occurred after training.

Figure 3: Estimated change in performance from day 8 to day 102 for the three groups

Discussion

- 40 In summary, the post-experiment questionnaire reliably reflected the content of the training. Trained interviewers acknowledged that the training improved their knowledge of the sampling and selection processes and helped them understand the situation they face. They did not feel that the training itself helped them transfer the acquired knowledge into practice, but they were more likely than high performers to feel their performance had improved. High performers were more likely to feel self-confident but did not feel more knowledgeable than trained interviewers.
- 41 Since the post-training questionnaire was distributed two weeks after the training, two biases may have occurred. One stems from the fact that trained interviewers may have discussed the information they received with the other interviewers, so that the training

also had an influence on the other interviewers. The other is that the delay between training and the distribution of the questionnaire could have introduced non-response bias, since some interviewers, with specific characteristics, left in the interval. However, because trainees tend to rate training very highly immediately after the training for a number of reasons (they are taken away from work, cared for, stimulated, etc.), it is important to determine what the trainees retained two weeks later.

- 42 Employee retention is an issue that should be followed up on in subsequent research. For private pollsters, it may constitute a by-product of training that is almost as important as improved performance, since high turnover is very costly, particularly when time and money has been put into training and when training is effective.
- 43 Finally, the results showed a substantial effect of training on performance. However, the effect was not sufficient to raise trained low performers to the level of high performers. The effect also tended to plateau with time. This was a first experiment. Improvement in the training could bring about a more substantial impact. It should be stressed, however, that it was possible to measure this effect using the NCPi as an index of performance, but not using cooperation rate at first contact, because the latter measure cannot be calculated when interviewers are working on appointments or refusal conversion.

Conclusion

- 44 The experimental one-hour training sessions appear to have had some significant effect on two of the three factors we were examining; *i.e.*, attitudes and performance. However, as with any experiment, one must ask whether a Hawthorne effect was at work, whereby merely giving interviewers attention creates a subsequent improvement in performance. While this could be the case, all the results point in the same direction and, if the experiment did induce a Hawthorne effect, this one had the advantage of placing relatively few demands on the organization and of being inexpensive.
- 45 It is also obvious that the experiment should be repeated and improved in order to seriously prove its impact. Improvements should, in our view, go in the following directions:
- 46 - The one-hour format should be retained as much as possible, but the content should be split into two one-hour sessions. This would allow for more interviewer input and questions during training.
- 47 - The use of M&Ms has many advantages. It allows for a lively and flexible presentation. However, it could be improved in order to systematically go through all the relevant information related to sampling and within-household selection. Training should also include a clear explanation of how phone numbers are generated: trainees asked this question frequently.
- 48 - The training related to prevention of refusals should be made more lively and should better encourage the transfer of knowledge into practice without resorting to a “drill and practice” approach (too long and not always very appreciated).
- 49 - Training should include an explanation of where the data are going; *i.e.*, into percentages of people doing such and such activity. This also implies that eating areas for interviewers should contain books, research reports, etc., which would give them a better

idea of their work's end product. These types of interventions convey to interviewers the message that they are important, knowledgeable employees.

BIBLIOGRAPHY

- Auster, Carol J. (2000). "Probability Sampling and Inferential Statistics. An Interactive Exercise Using M&M's". *Teaching Sociology*, Vol. 28, pp. 379-385.
- Carton, Ann (2000). "An Interviewer Network: Constructing a Procedure to Evaluate Interviewers". *Bulletin de Méthodologie Sociologique*, Vol. 67, pp. 42-53.
- Cantor, David, Allen, Bruce, Schneider, Sid J., Hagerty-Heller, Tracy and Yuan, Angela (2004). *Testing an Automated Refusal Avoidance Training Methodology*. American Association for Public Opinion Research (AAPOR), 59th annual conference, May13-16, 2004.
- Dijkstra, Wil and Smit, Johannes H. (2002). "Persuading Reluctant Recipients in Telephone Surveys". In editors Groves, Robert M., Dillman, Don. A., Eltinge, John. L. and Little, Roderick J. A., *Survey Nonresponse*. New York: John Wiley and Sons, Inc., pp. 121-134.
- Downey-Sargent, Kathryn and O'Hare, Barbara C. (2004). *A Pre and Post-Test Experiment Investigating the Effectiveness of Voice Training for Telephone Interviewers*. American Association for Public Opinion Research (AAPOR), 59th annual conference, May13-16, 2004.
- Durand (2005). "Measuring Interviewer Performance in Telephone Surveys". *Quality and Quantity*, 39, pp. 763-778
- Goyder, John (1988). *The Silent Minority*. United Kingdom: Oxford University Press.
- Groves, Robert. M. and McGonagle, Katherine. A. (2001). "A Theory-Guided Interviewer Training Protocol Regarding Survey Participation". *Journal of Official Statistics*, Vol.17, no 2, pp.249-265.
- Hox, Joop (2002). *Multilevel analysis*, London: Lawrence Erlbaum Associates.
- Lemay, Michael and Durand, Claire (2002). "The Effect of Interviewer Attitude on Survey Cooperation". *Bulletin de Méthodologie Sociologique*, No. 76, pp. 27-44.
- Mayer, Thomas S. and O'Brien, Eileen (2001). *Interviewer Refusal Aversion Training to Increase Survey Participation*. Proceedings of the Annual Meeting of the American Statistical Association, August 5-9, 2001.
- McConaghy, Mark and Carey, Siobhan (2004). *Training to Help Interviewers Avoid Refusals: Results of Stage 1 of a Pilot Using Avoiding Refusal Training (ART) with Interviewers at the Office of National Statistics, UK*. American Association for Public Opinion Research (AAPOR), 59th annual conference, May 13-16, 2004, Phoenix, Arizona.
- McConaghy, Mark and Carey, Siobhan (2005). *What happens after ART? Results of two experiments designed to improve response rates with interviewers at the Office of National Statistics, UK*. American Association for Public Opinion Research (AAPOR), 60th annual conference, May 12-15, 2005, Miami, Florida.
- Nagin, Daniel. S. (1999). "Analyzing Developmental Trajectories: A Semiparametric, Group-Based Approach". *Psychological Methods*, Vol. 4, No. 2, pp. 139-157.

- O'Brien, Eileen, Mayer, Thomas S., Groves, Robert M. and O'Neill, Grace E. (2002). *Interviewer Training to Increase Survey Participation*. Proceedings of the Annual Meetings of the American Statistical Association, August 11-15, 2002.
- Raudenbush, Stephen W. and Bryk, Anthony S. (2002). *Hierarchical Linear Models*, London: Sage.
- Singer, Judith D. and Willett, John B. (2003). *Applied Longitudinal Data Analysis*, Toronto: Oxford University Press.
- Snijders, Tom and Boskers, Roel (2000). *Multilevel Analysis*. London: Sage.
- Van der Zouwen, Johannes, Dijkstra, Wil and Smit, Joannes H. (1991). "Studying Respondent-Interviewer Interaction: The Relationship Between Interviewing Style, Interviewer Behavior and Response Behavior". In editors Biemer, Paul P., Groves, Robert M., Lyberg, Lars E., Mathiowetz, Nancy A. and Sudman, Seymour, *Measurement Errors in Surveys*. New York: John Wiley and Sons, Inc., pp.419-437.

NOTES

1. Montreal has a modest English-speaking population but a fair proportion of bilingual people with French as their mother tongue. This makes it difficult to recruit English-speaking interviewers and interviewers who speak English with an accent that is not too pronounced.
2. This is, the file produced by the CATI software that gives the basic information on all the calls performed: date and time of call, duration, result and ID of the interviewer performing the call.
3. A value of one on the NCPI is equivalent to having completed interviews of average length (24.6 minutes) during the whole work time.
4. The analysis according to day in the field rather than interviewer experience on the project was used for selection purposes since it evaluated the performance of interviewers at the same moment, whatever their experience on the project.
5. Measured by an additive scale of five related items (Cronbach alpha=.81).
6. Measured by an additive scale of the two items (Cronbach alpha=.58).
7. A variable that has significant random effects does not act in the same way for different interviewers.
8. Interaction effects of group membership on both the linear and quadratic effects of time were tested in order to assess whether the basic evolution of performance differed by group. These effects were either not, or did not remain, significant when the effect of training was entered later on. We do not present these models.
9. This variable has a value of zero for all interviewers before training and increments from 1 to 45 (end of field work) only for the trained interviewers.
10. Therefore, the equations for the final model are:
Level-1 Model:
$$NCPI_i = B_0 + B_1*(TIME) + B_2*(TIME^2) + B_3*(TIME - post-training) + B_4*(TIME^2 - post-training) + R$$

Level-2 Model:
$$B_0 = G_{00} + G_{01}*(PRETRAINING\ GROUP) + U_0$$

$$B_1 = G_{10} + U_1$$

$$B_2 = G_{20}$$

$$B_3 = G_{30}$$

$$B_4 = G_{40}$$

Final Model:
For the high performers:
$$NCPI_i = (-.77178 + .44519) + .01462*(TIME) - .00005*(TIME^2)$$

For the Control group:

$$NCPI = -.77178 + .01462 * (TIME) - .00005 * (TIME^2)$$

For the trained group:

$$NCPI = -.77178 + .01462 * (TIME) - .00005 * (TIME^2) + .02531 * (TIME - \text{post-training}) - .0005 * (TIME^2 - \text{post-training}) + R$$

ABSTRACTS

This article presents an experiment in which a short cognitive training session was given at the private pollster conducting the 2004 Canadian Addiction Survey of 14,000 people. Overall, 79 interviewers worked on the project for a total of 2,436 interviewer-days, interviewers worked a maximum of 65 days, and the survey was in the field for 109 days. We devised a one-hour training session that focused on knowledge acquisition in two areas: sampling and selection within household, and reasons for refusals. The evolution of interviewers' daily performance, analyzed using group-based trajectory analysis, allowed us to identify low (n=25) and high (n=42) performers two weeks before training. The training session was aimed at low performers and newly hired interviewers. Among these, all the interviewers working on training day (n=18) attended one of the three training sessions. The impact of the training was assessed in three areas: knowledge acquisition and attitudes (through a post-training questionnaire two weeks later), retention, and performance. The results show that trained interviewers a) felt that the training helped them understand sampling and the reasons for refusals b) were more likely than high performers to feel that their performance had improved since the beginning of the project, c) were likely to go on working on the project after training longer than the control group, and d) improved their performance by .04 per day after training, compared with .015 per day before training, almost reaching the performance of high performers.

Un examen de l'efficacité de la formation complémentaire pour les interviewers d'enquêtes téléphoniques : Cet article présente une expérience faite à l'agence d'enquêtes responsable pour l'enquête canadienne de 2004 sur la consommation de drogues (N=14.000) après une courte séance de formation cognitive. Un total de 79 intervieweurs y ont participé représentant 2.436 journées-intervieweur de travail, avec des intervieweurs travaillant un maximum de 65 jours : l'enquête de terrain a duré 109 jours. Nous avons conçu une séance d'une heure centrée sur l'acquisition des connaissances dans deux secteurs : l'échantillonnage et la sélection dans le foyer d'une part, et les raisons de refus d'autre part. L'évolution quotidienne des performances des intervieweurs, dont les trajectoires basées sur des groupes ont été analysées, nous ont permis d'identifier les intervieweurs à performances basses (n=25) et hautes (n=42) deux semaines avant la formation. La séance de formation était à l'intention des intervieweurs à performances basses ou des nouveaux intervieweurs. Parmi ces intervieweurs, tous ceux qui travaillaient le jour de la séance (n=18) ont participé à l'une des trois séances de formation. Les retombées de la formation ont été évaluées dans trois secteurs : connaissances acquises et attitudes (via une questionnaire rempli deux semaines plus tard), rétention de connaissances et performances. Les résultats montrent pour les intervieweurs participants (a) qu'ils trouvaient que la formation les avait aidés à comprendre l'échantillonnage et les raisons de refus, (b) qu'ils sentaient plus que les intervieweurs à hautes performances que leur performances s'étaient améliorées depuis le début du projet, (c) qu'ils avaient tendance à continuer plus longtemps à travailler sur les projet qu'un

groupe de contrôle, et (d) qu'ils ont amélioré leur performance de .04 par jour après la formation, par rapport à .015 par jour avant la formation, atteignant presque la performance des intervieweurs à hautes performance

INDEX

Keywords: Growth Model, Interviewers, Multilevel Analysis, Performance, Retention, Survey, Training

Mots-clés: Analyse multi-niveaux, Enquêtes, Formation, Intervieweurs, Modèle de croissance, Performance, Rétention de connaissances

AUTHORS

CLAIRE DURAND

Dept. of Sociology, Université de Montréal; Claire.Durand@umontreal.ca