An Experiment in Cognitive Training of Telephone Survey Interviewers

Paper submitted by

Claire Durand
Marie-Eve Gagnon
Christine Doucet
Eric Lacourse

Do not quote without permission

Correspondence to:
Claire Durand
Claire.Durand@umontreal.ca
Department of sociology,
Université de Montreal,
C.P. 6128, succ. Centre ville,
Montreal (Quebec), Canada H3C 3J7

An Experiment in Cognitive Training of Telephone Survey Interviewers

Introduction

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.

The training of interviewers working in survey research often takes place in an employment context where incentives for training are few. 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.

In the context of social research, interviewers are asked to carry out an activity that is considered to be scientific in nature—a job once reserved to specialists. However, there are no credentials that may testify to an interviewer's competence, nor is the job considered a career. In addition, in private firms, there is a trade-off to be made between quality and cost effective production, between the need for competent but flexible and not too expensive employees, keeping in mind the costs associated with recruitment and training. It is against this background that the research team devised a training session focusing on improving cooperation rates in order to complement the initial training provided by a private survey firm specialized in social research.

The context: The training of interviewers

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, Mayer, Groves and O'Neill, 2001; McConaghy and Carey, 2004; Cantor et al., 2004) have experimented with training along the lines presented by Groves and McGonagle (2001) and inspired from Groves and Cooper (1998). 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, 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-training evaluation, personality tests, post-training evaluation, observation scores, etc.) to better evaluate what was learned during training and relate it to subsequent performance.

The two experiments presented by Groves & 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 report a 14-percent improvement in cooperation following training; however, this increase is not clearly related to performance in training. They also conclude that "we are 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."

However, Mayer & O'Brien (2001) did exactly that, i.e., they followed up on Groves & 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, who received no training; a "before" group, who received the eight-hour training before the beginning of the project (two two-week surveys); and a

"between" group, who received the training 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 to 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., 2001) 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.

McConaghy and Carey (2004) 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. They had similar results: some improvement in response rates (around 9 percent) but some difficulty in assessing the magnitude of the change that may have been due to training itself. In addition, the low number of interviewers (n=15) precluded the possibility of carrying out in-depth statistical analyses. 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.

In summary, authors using the ART program sometimes found an 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. The possibility of a Hawthorne effect is present. Only one of the experiments was conducted during a RDD phone survey, and in most experiments, the pool of interviewers was rather homogenous, experienced, and small. Finally, when asked (Groves and McGonagle, 2001), interviewers found the training useful but did not enjoy it.

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 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."

Dikstra & Smith (2002) recorded interactions in an attempt to understand what works and what does not. The interviewers were all women, students in a department of communication, with no experience as interviewers; the generalizability of the data is thus 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. Van der Zouwen, Dikstra & Smith (1991) reported that when interviewers are trained to adopt a more formal style, they tend with time to return to a more personal style.

Summarizing these various results from experiments, we might conclude that the impact of training appears to be generally low. This may be due to the content of the training and the methods used. It may also be due to an inappropriate setting (i.e., 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 --, length of project – too short to enable selection of interviewers, training and assessment of impact). In addition, the measure of performance generally used, the cooperation rate at first contact, may be unreliable, thus preventing identification of an effect. 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 one does or does not have and that it cannot be improved by training, it appeared necessary to tackle these four possible reasons for the low impact of training, i.e., the setting, the content and methods used as well as the measurement and analysis of performance.

The experiment

The setting

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, 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) is 47% for a total of 13999 completed interviews. The average length of interviews is 24.6 minutes.

Overall, 79 interviewers worked on the project for a total of 2436 interviewer-days; interviewers worked a maximum of 65 days and the survey was in the field for 109 days. Since the firm is usually conducting 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 completed the questionnaire, and 54 of them gave their ID codes, allowing us to match their questionnaire with performance data.

Table 4 in Appendix 1 gives 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 exact 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%) consider themselves as regular employees; they split almost equally between full time and part-time. More than a third (36%) worked 25 hours or less per week while one on four (23%) worked more than 35 hours a

week.

Content and methods

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. The assumption is that knowledge will translate into the development of abilities because interviewers will better understand why they are asked to do within household selection, to convince selected respondents and to perform refusal conversion and what happens when they try to carry out these tasks. The goal is to increase interviewers' self-confidence and intrinsic motivation in order to help them find their own solutions to the problems they face, i.e., to tailor according to their own personal style of interaction. We postulate that tailoring has to do not only with respondents but also with interviewers' habits in dealing with people. If their reaction to respondents seemed artificial or pre-programmed, it may not work.

Two types of information were identified as lacking and possibly helpful to interviewers. Concrete information on sampling in general and on the particular sampling frame used for surveys of the general population, should improve interviewers' understanding of why they had to select respondents and convince those selected to answer the survey. The second type of information, pertaining to why people refuse to participate in surveys and how experienced interviewers deal with refusals, allows for a better understanding of the social interaction and of what may be happening in the household they are calling. The stated goal is to reduce the stress associated with "taking it personally," i.e., taking too much responsibility for the refusals and therefore becoming paralyzed by stress.

One feature to consider in setting up a training session is 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.

[1] Montreal has a modest English-speaking population and a fair proportion of bilingual people with French as their mother tongue. This makes it more difficult to recruit English-speaking interviewers or interviewers who speak

The second decision concerns a feasible setup for the training sessions. It was decided to have short training sessions. In the context of a private firm's activities, it seemed more realistic to devise a short session that would be easier to implement in various situations. In addition, a short session has the advantage that it is 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 acted as a control group. 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-up on appointments and refusal conversion).

The training itself

The first step in the process is the identification of the low performers. During the first weeks of the fieldwork, the administrative database[3] was collected every two weeks in order to monitor the evolution of performance. The performance itself was measured using an index of the net contribution to performance (NCPI [4]: Durand, accepted). This index, contrary to the cooperation rate at first contact, takes into account all the tasks performed by interviewers, including follow-ups on appointments and refusal conversion. Therefore, it can be computed whatever the task performed. Semi-parametric group-based trajectory analysis (Nagin, 1999) of the evolution of daily performance was used in order to identify two groups, which may be termed high and low performers. Figure 1 illustrates the evolution of performance from day 7 in the field (after the Christmas break) up until two weeks before training, at the moment where interviewers were selected. The analysis identifies 25 interviewers in the low performance trajectory and 42 in the high performance. The trajectories of low versus high performers differ in the average starting level (-.87 compared to -.40). The average rate of improvement is similar (.0159 per day in the field) but the average daily performance of the high performers appears to become more stable with time.

English with an accent that is not too pronounced.

[3] i.e., the file produced by the cati software which gives the basic information on all the calls performed: date and time of call, duration, result and Id of the interviewer performing the call.

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 on who to train.[5] In addition, 10 interviewers were hired between the identification phase and the training and added up to the group to be trained. Table 1 in Appendix 1. shows that there is little difference in the characteristics 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 the trained interviewers were more likely to be men.

All the identified low performers and the 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. They 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.

The first part of the training pertained to sampling. To liven up the training, M&M bags were used (see: Auster, 2000) in order to explain concretely different notions: 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. 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 refusers (mostly from

---

[4] A value of one on the NCPI is equivalent to having completed interviews of average length (24.6 minutes) during the whole work time.

[5] The analysis according to day in the field rather than interviewers' experience with the project was used for selection purposes since it allowed to evaluate the performance of interviewers at the same moment, whatever their experience on the project. The classification of interviewers does not differ much between the two analyses and it almost does not vary between the identification date and the day before training. The 10 interviewers hired during these two weeks split half between high and low performers on the day before training. X of the high performers and y of the low performers ended in the training group.

Goyder (1988), *The Silent Minority*) and information gleaned from interviews of high performers on how to deal with refusals.

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 old woman if she has ever smoked marijuana? We answered these questions, sticking to the analogy of bags of M&Ms 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 will not be correct if the non-smokers are not interviewed. As for the second part, some interviewers seemed very surprised by part of the presentation, saying that the information was going 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 at very least a "critical success."

Results: Impact of training

Training may 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

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 how much they understand sampling and within-household selection. Version B, distributed only to the trained interviewers, had one more section concerning the evaluation of the training per se i.e., whether the training helped 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).

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 2 shows the distribution of answers to questions pertaining to interviewer related items of the questionnaire and self-reported 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 vary between 48% (20 of 42 feel they control the situation during interviews) and 60% (26 of 43 fully agree that people accept because the interviewer is self-confident). In addition, 33% (14 of 43) say they are capable of finding arguments most of the time and 60% that the role of the interviewer in the decision to participate is very important. A scale of these items can be computed[9]. When the trained interviewers are compared to the high performers group – low untrained performers being

[9] Measured by an additive scale of five related items (Cronbach alpha=,81)

too few for a significant comparison to take place – it appears that the high performers are more likely to be self confident and to attribute cooperation to their own work (F=6.23, p=.017) than the trained interviewers.

Finally, 49% (21 of 43) say that they know very well how telephone numbers are selected and 81% (35 of 43) know very well why selection within household is done. Analysis shows that high performers and trained interviewers do not differ in their self-reported knowledge[10] of sampling and selection (F=.446, p=.51).

Finally, trained interviewers are more likely than high performers to declare that their self-confidence has improved since the beginning of their work on the project (Lr test=6.47, p=.039). All the trained interviewers say their capacity to convince has improved much (7 of 13) or somewhat (6) while 19 of 26 high performers say their performance has improved, the rest (7) stating that it has remained stable.

b) Employee retention

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.

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 control group, only seven of 17 were still working on the training date: ten had left between the

[10] Measured by an additive scale of the two items(Cronbach alpha=,58).

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 significant (F=1.35, p=.27).

c) Performance

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 is necessary to have a reliable measure of performance and to examine the evolution of performance throughout the project, allowing to estimate how the performance has evolved after the training session. Performance may be seen as nested within interviewers so that multilevel growth models may be used.

Multilevel longitudinal analysis is a classical approach to analyze this type of data (Singer & Willett, 2003; Hox, 2002; Raudenbush and Bryk, 2002; Snijders and Boskers, 2000). The analysis does not use the data for the two first days of work of each interviewer since performance is not reliable enough during these first learning days. The first thing to decide was whether time itself (day in the field) or each interviewer's experience in the project would be used as the time variable. It was decided to use real time since it is the same for all subjects at the same moment. Table 4 shows the results for the analyses performed in order to identify a parsimonious and theoretically sound model. In the current situation, performance can improve only to a certain level, an optimal level in a given situation. When this level is reached, performance should normally plateau. Therefore, it is sound to hypothesize the presence of a quadratic component in the effect of time. Model 0 includes only the linear effect of time. It shows that there is as much variance in performance between interviewers (random variance of the intercept = .09368) than 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. In model 1, the quadratic effect of time is added. This effect is also significant but does not vary between interviewers. Since the variance within

interviewers has increased to .09818, this model will be considered the base model against which the following models will be compared. In model 2, membership in the trajectories used for selection is entered in order to control for the pre-training base situation. The between interviewers' variance decreases from .09818 to .01996 which means that the categorization explains 79.7% of the variance between interviewers. The intercept- i.e., the starting performance – of the high performers is .466 higher than that of the low performers. The effect of group membership was tested on the linear and quadratic effect of time in order to see if the evolution of performance differs with group membership. These effects are not or do not remain significant when the effect of training is entered.

The following models present the effect of training. This effect can be modeled is different ways (see recommendations by Singer and Willett, 2003) . Training may have an effect on the intercept – a jump in performance following training – or on the post-training evolution of performance. Multiple tests have directed us to the final model presented. In Model 3 , the linear effect of time after training is entered. This variable has a value of zero for all interviewers before training and takes an incrementing value from 1 to 45 (end of field work) only for the trained interviewers. Model 3 shows that the net effect of time on post training performance is .00789 (p=.011). Model 4 shows a slightly different story. The quadratic effect of time 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. Final tests of the significance of the random effects bring us to the final model where only the intercept and the general linear effect of time have random effects that are significant. This means that there is still some variation in the average performance of interviewers and in the average effect of time that is not explained. Figure 2 illustrates the results of the average performance of the three groups, i.e., the high performers, the low performers and newly hired untrained interviewers and the trained interviewers, and this from day 7 (after the Christmas break) to day 100 (nine days before the end of fieldwork), the last days not being typical (not many interviewers working, tasks left more specific, etc.)

Discussion

In summary, the post experiment questionnaire taps reliably the content of the training. Trained interviewers acknowledge that the training improved their knowledge of the sampling and selection process and helped them understand the situation they face. They do not feel that the training itself helped them transfer the acquired knowledge into practice. High performers are more likely to feel self-confident but do not feel more knowledgeable than trained interviewers.

Since the post-training questionnaire was distributed two weeks after the training, two biases may have occurred. One comes from the possibility that trained interviewers discussed the information they received with the other interviewers, so that the training also had an influence on the other interviewers. Another is that the delay between the training and the distribution of the questionnaire introduced non-response bias since some interviewers, with specific characteristics, had left in the mean time. However, since trainees tend to give very good ratings to training immediately after the training for a number of reasons (they are taken away from work, cared for, stimulated, etc.), it is important to know what remains of the training in trainees' minds two weeks later.

Retention is an issue that should be followed in subsequent research. For private pollsters, it may be an effect of training that is almost as important as performance since high turnover is very costly.

Finally, the results show a substantial effect of training on performance. However, this effect is not sufficient to bring the trained low performers to the level of the high performers. It also tend to plateau with time. This was a first experiment. Improvement in the training could bring a more substantial impact. It has to be stressed however that this effect could be measured using the NCPi as an index of performance. It could not be measured using the cooperation rate at first contact because this measure cannot be computed when interviewers are working on appointments or refusal conversion.

Conclusion

The experimental one-hour training sessions appear to have had some significant effect on two of the three different factors we were examining i.e., employee retention and performance. However, as with any experiment, one must ask whether a Hawthorne effect is at work, where the mere attention paid to interviewers is the real explanation for the 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 being relatively undemanding and inexpensive for the organization.

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:

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.

- 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 where the phone numbers come from, how they are generated: this was a reiterated request of interviewers.
- The training related to prevention of refusals must be made more lively and should help better the transfer of knowledge into practice without resorting to a “drill and practice” approach (too long and not very appreciated).
- Training should include an explanation of where the data is going, i.e., into percentages of people doing such and such activity. This also implies that dining rooms for interviewers should contain books, research reports, etc., which would give them a better idea of their work’s end product.

References

Auster, C. J. (2000). Probability Sampling and Inferential Statistics. An Interactive Exercise Using M&M's. *Teaching Sociology*, Vol. 28, p. 379-385.

Billiet, J. & Loosveldt, G. (1988). Improvement of the quality of responses to factual survey questions by interviewer training. *Public Opinion Quarterly*, Vol. 52, p.190-211.

Carton, A. (2000). An Interviewer Network : Constructing a Procedure to Evaluate Interviewers. *BMS (Bulletin de Méthodologie Sociologique)*, Vol. 67, p. 42-53.

Cantor, D., Allen, B., Schneider, S. J., Hagerty-Heller, T. & Yuan, A. (2004). *Testing an Automated Refusal Avoidance Training Methodology*. American Association for Public Opinion Research (AAPOR), 59th annual conference, May13-16, 2004.

Dijkstra, W. & Smit, J. H. (2002). Persuading Reluctant Recipients in Telephone Surveys. In Groves, R. M., Dillman, D. A., Eltinge, J. L. & Little, R. J. A. (Eds), *Survey Nonresponse*. New York: John Wiley & Sons, Inc., p. 121-134.

Downey-Sargent, K. & O'Hare, B. 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, C. (accepted). How to Measure Interviewer Performance in Telephone Surveys. *Quality and Quantity*, 2004, 25 pages.

Goyder, J. (1988). *The Silent Minority*. United Kingdom: Oxford University Press.

Groves, R. M. & Couper, M. P. (1998). *Nonresponse in Household Interview Surveys*. New York: John Wiley.

Groves, R. M. & McGonagle, K. A. (2001). A Theory-Guided Interviewer Training Protocol Regarding Survey Participation. *Journal of Official Statistics*, Vol.17, no 2, p.249-265.

Hox, J. (2002). *Multilevel analysis*, London: Lawrence Erlbaum Associates, 304p.

Lemay, M. & Durand, C. (2002). The Effect of Interviewer Attitude on Survey Cooperation. *BMS (Bulletin de Méthodologie Sociologique)*, No. 76, p. 27-44.

Mayer, T. S. & O'Brien, E. (2001). *Interviewer Refusal Aversion Training to Increase Survey Participation*. Proceedings of the Annual Meeting of the American Statistical Association, August 5-9, 2001.

McConaghy, M. & Carey, S. (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, May13-16, 2004.

Nagin, D. S. (1999). Analyzing Developmental Trajectories: A Semiparametric, Group-Based Approach. *Psychological Methods*, Vol. 4, No. 2, p. 139-157.

O'Brien, E., Mayer, T. S., Groves, R. M. & O'Neill, G. E. (2002). *Interviewer Training to Increase Survey Participation*. Proceedings of the Annual Meetings of the American Statistical Association, August 11-15, 2002.

Raudenbush, S.W. and A.S. Bryk (2002). *Hierarchical Linear Models*, London : Sage, 485p.

Singer, J. D. and J.B. Willett, (2003). *Applied Longitudinal Data Analysis*
Snijders, T. And R. Boskers (2000). *Multilevel Analysis*. London : Sage, 266p.

Van der Zouwen, J., Dijkstra, W. & Smit, J. H. (1991). Studying Respondent-Interviewer Interaction: The Relationship Between Interviewing Style, Interviewer Behavior and Response Behavior. In Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A. & Sudman, S. (Eds), *Measurement Errors in Surveys*. New York: John Wiley & Sons, Inc., p.419-437.

Table 1 Evaluation of the training session

	Total	Fully agree	Somewhat agree	Fully or somewhat disagree
	N	N	N	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

Table 2 Interviewers self-perception of their role, behavior 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
			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

Table 3
Results of longitudinal multilevel analysis of evolution of performance

	Model 0 linear	**Model 1 quadrat**	**Model 2 "+ group effect**	**Model 3 "+ post training linear**	**Model 4 "+ post training quadratic**	**Final model Without non-sig. Random**
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 lin				.00005*	.00019*	"-----
post train. Time quad.					.00000	"-----
Within interv. 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)

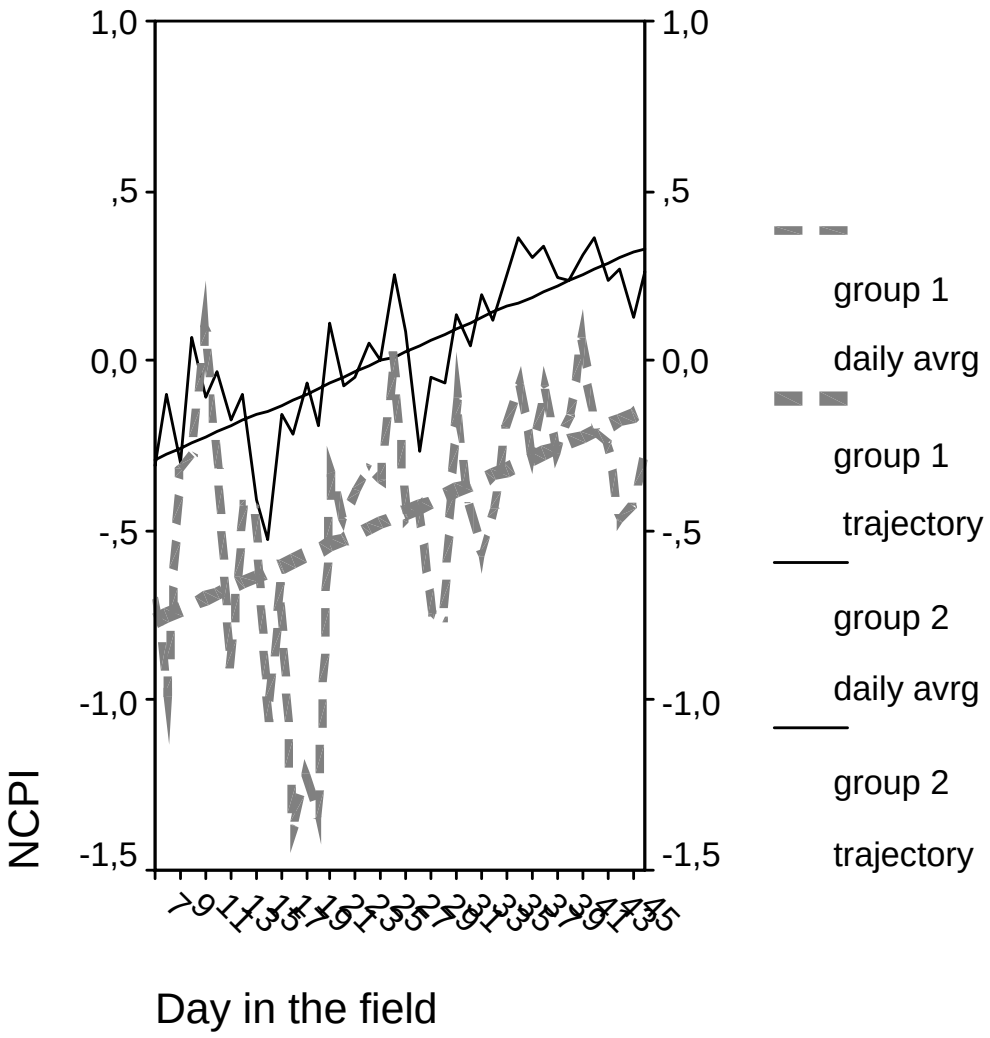

Figure 1. Trajectories of low and high performers on Feb. 18, two weeks before training.

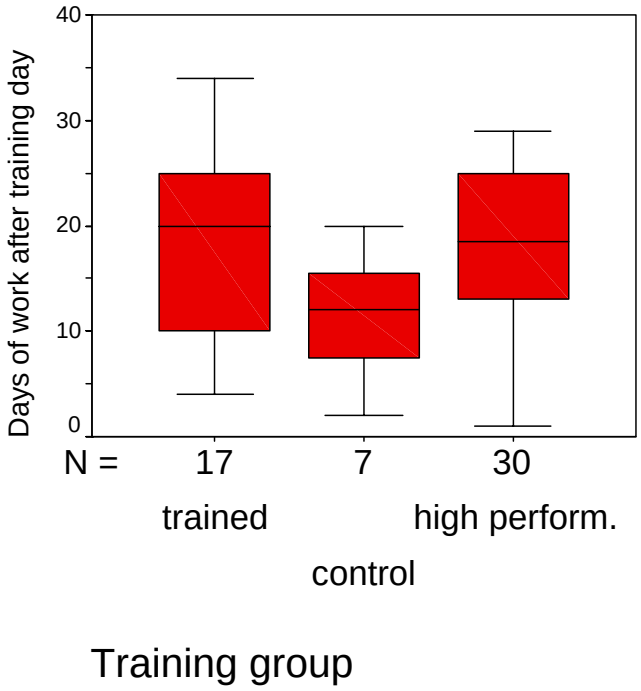

Figure 2. Employee retention according to training group

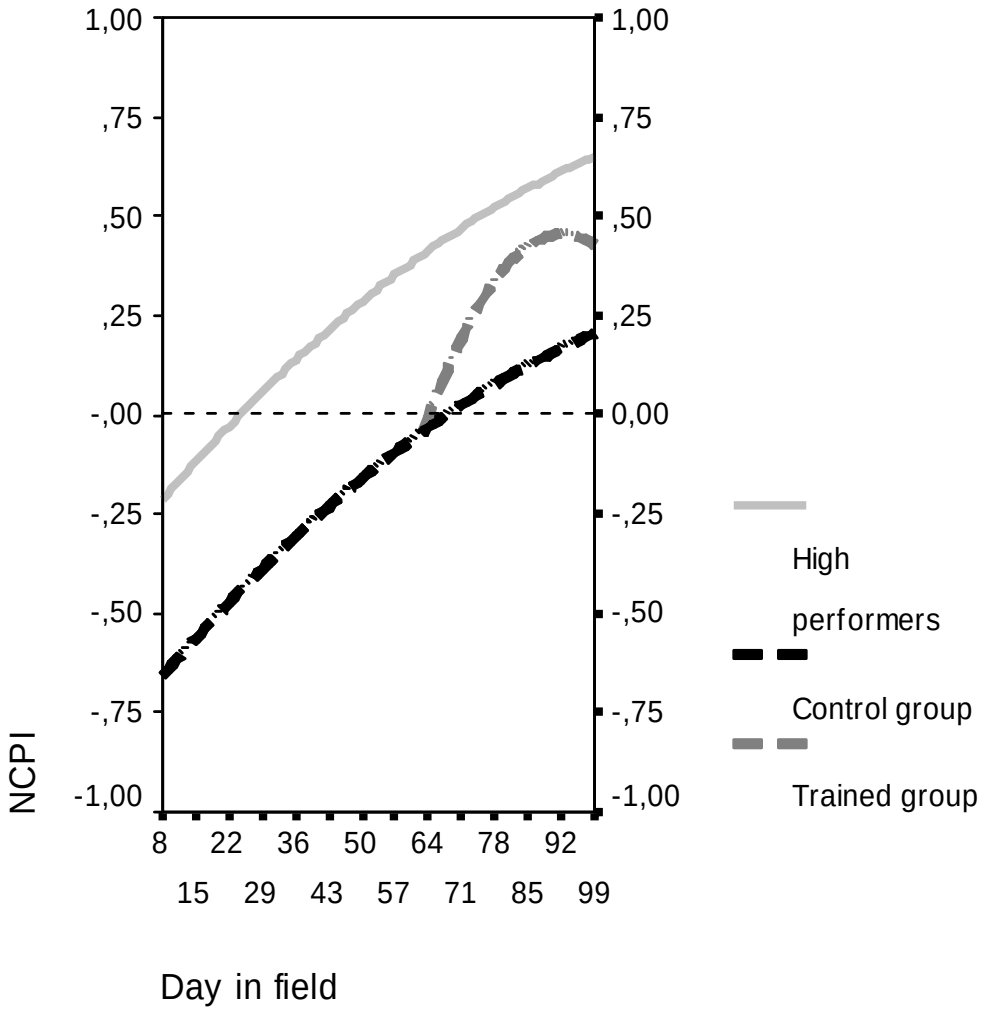

Figure 3. Estimated evolution of performance from day 8 to day 99 for the three groups

Appendix 1
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**	75%	75%	80%	74%	74%	79%
	Other	25%	26%	20%	27%	26%	21%
	N	**100%**	**8**	**15**	**34**	**43**	**14**
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%
	N	**100%**	**8**	**13**	**34**	**42**	**13**
Experience as interviewer	**0-3 months**	25%	13%	20%	29%	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	**Complementar**	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**