

# A COGNITIVE AND LINGUISTIC EXPLORATION OF THE MEANING OF TRAINING: IMPLICATIONS FOR SURVEY DESIGN<sup>1</sup>

Pamela Campanelli (SCPR) and Joanna Channell (Univ. of Birmingham) with  
Liz McAulay (Univ. of Birmingham), Antoinette Renouf (Univ. of Liverpool), and Roger Thomas (SCPR)  
Pamela Campanelli, SCPR, 35 Northampton Square, London EC1V 0AX England

**Keywords:** Surveys, training, cognitive methods, applied linguistics, questionnaire design

## 1. Training: A Problem of Under-Reporting

It is common knowledge that survey respondents may not always interpret words and concepts used in surveys in the way researchers intend (cf Belson 1981 among many). The vast array of question miscommunication outcomes can be roughly simplified into two broad classes of discrepancies: those cases where respondents have a narrower interpretation than researchers and those where respondents have a wider interpretation than researchers (Campanelli, Martin and Rothgeb 1991). Cases where respondents' definitions are broader than researchers' should ideally be handled by interviewers; or information should be collected so that analysts can exclude certain reports at a later phase. More problematic is the situation where respondents employ a narrow interpretation. In this case, neither the interviewer nor the analyst has any way of knowing that an activity which the analyst would consider relevant, has not been reported.

In the UK there has been recent concern about the conceptual ambiguity which surrounds the term *training* (Robson 1992). More specifically the UK Employment Department (ED) has suspected:

- 1 that the general population uses the term *training* to refer to a much narrower set of activities than those understood by training professionals;
- 2 that for most people *training* is that which happens in formal courses and which leads to qualifications;
- 3 that employees have a narrower understanding than employers;
- 4 that activities included in the definition of *training* will vary with different occupational groups, age groups, and genders;
- 5 that activities which are self-initiated and/or self-funded are less likely to be included than those which are initiated and/or funded by an employer, and
- 6 that there is a fuzzy boundary between training and education for most people.

To the extent that these assumptions are true, a serious amount of under-reporting of training

activities may take place in surveys. Assumption 4 deserves special comment, as the under-reporting bias is correlated with major survey explanatory variables. Such differential rates of under-reporting could seriously bias comparisons of training experiences between age cohorts, men and women, and different occupational groups.

## 2. The Meaning of *Training* Project

It was in this context that in 1993 the Survey Methods Centre at SCPR (London, UK) in conjunction with the Research and Development Unit for English Studies (then at the University of Birmingham, UK) were commissioned by the UK Employment Department to explore the validity of their assumptions with respect to respondents' understanding of *training* and the resultant implications for survey design. Our objectives were:

- a) to investigate people's understanding of the concept of *training*;
- b) to investigate people's use of the word *training* and its associated word forms (*train*, *trains*, *trained*, *trainer*, and *trainee*);
- c) to investigate how speakers talk/write about the range of activities which could be identified as *training* when they do not use the word itself, &
- d) to devise and pretest a revised form of questioning for surveys.

We tackled objectives (a) - (c) with two basic strategies: a linguistic and a cognitive approach. With the linguistic approach we studied the meaning of the word *training* as a product of the situations and linguistic contexts where it is used. The linguistic approach also investigated the ways in which people refer to the category 'training' when they do not use the word *training* itself. For example, a person saying *I did a medical degree* or *I went on a one-day course* is referring to activities which are members of the cognitive category *training*.

The cognitive approach looked at the relationship between the term *training* and the cognitive category to which it refers. As with many words, the membership of the category consists of some central and typical examples, some peripheral examples and some fuzzy examples which might or might not be members. The category can also partly be defined by reference to items which are definitely

not members (for work on cognition and categorisation see Rosch 1975; Rosch and Lloyd 1978; Neisser 1987). In the case of *training*, a central and typical case could be doing an employer-sponsored course, a peripheral item could be doing a self-funded course, and a fuzzy item could be reading a manual. An activity which clearly is not *training* in this context is, say, doing your laundry.

Given the background of a combination of cognitive and linguistic issues, five specialised techniques were used. Three of these were from the survey-based tradition and drawn from the large variety of question testing techniques which survey researchers now have available to them for diagnosing problems of question meaning (cf Forsyth and Lessler 1991). These included cognitive interviewing methods, vignette analysis, and focus group discussions. Each of these can be used to determine how ordinary people categorise various activities which might be included in a wider definition of 'training' and are described more fully in Section 3. The other techniques used were from the language-based tradition and included corpus linguistics and conversation analysis. Corpus linguistics looks at the patterns of occurrence of words and expressions in a large database of naturally occurring language (for a description see Sinclair 1991) while conversation analysis focuses on the goals and problems displayed through language by speakers or writers in selected stretches of conversation or text (see Drew and Heritage 1992). Both corpus linguistics and conversation analysis study language in the way it is actually used, rather than the way people think they use it. These techniques therefore complement the cognitively directed survey work by looking at language behaviour about training in natural, non-experimental situations.

This document provides a few highlights of the survey-based exploration of the meaning of training. A full discussion of both the linguistic and survey approaches is contained in a forthcoming UK Employment Department report by Campanelli *et al* (1994).

### 3. The Survey Exploration of Cognitive Issues

Cognitive Interviewing Methods. Cognitive interviews (NCHS 1989a) make use of a series of specific techniques which aim to probe how respondents go through the process of answering a given survey question. The main cognitive interviewing technique of relevance to the first phase of the project was a card sort application. Card sorting procedures (NCHS 1989b) are used to explore the natural taxonomies into which people

group lists of items. For the current research, a dimensional sort was used to determine what respondents do and do not consider to be *training*. To do this we developed 30 card sort items with the goal of evoking a broad range of response, with some items clearly being members or non-members of the category 'training' and the majority of the items being 'peripheral' or 'fuzzy' members of the category. These descriptions actually resemble vignette examples (the second technique) rather than typical card sort items (see Forsyth and Lessler 1991). This was done for two reasons. First, because of the complexity of the concept of *training* it was thought that it would be better to give respondents 'real life' situations with which they could identify, rather than abstract examples. Second, by having the card sort items resemble vignettes, the card sort task could be directly used as a pretest for the design of the vignettes.

Vignette Analysis. Vignette analysis was originally derived from the factorial survey approach described by Rossi and Anderson (1982) and is considered a good strategy with which to measure social judgements. It has been adapted as a tool to explore respondents' understanding of key survey concepts by embedding vignette examples in a respondent debriefing which is conducted as a supplement to a large-scale survey (see Martin *et al* 1986; Campanelli *et al* 1991). The project team commissioned the insertion of 12 vignettes into BMRB International's weekly Omnibus survey, covering a national probability sample of Great Britain during the 6th - 19th of May, 1994. This cost effective representative sample of 2,000 individuals provided the opportunity to conduct multivariate analyses to test the assumptions and served as an external validity check on the small-scale laboratory work.

A potential problem with vignettes is that respondents may react to different elements of the vignette from the ones under study, e.g. the age of the protagonist in the vignette rather than what the protagonist does. One generally studies the reasons for such variation using a factorial approach where an excessively large number of different vignette wordings are experimentally varied. When survey costs and implementational complexities prevent such a design, a qualitative approach can be used in which respondents are asked to discuss the reasons for their decisions. The qualitative approach of course assumes that respondents are conscious of the reasons for their decisions and that these can be verbalised.

**Focus Group Discussions.** We used a focus group setting to provide a qualitative framework for the exploration of people's responses to the self-administered card sort items. This is a novel approach which we found to be highly effective. For example, we had focus group respondents do the card sort task on an individual basis during what are often awkward moments before a focus group starts. Their answers were then individually recorded by the group leader. These tallied results later served as part of the topic guide for the group discussions as they informed the leader of which card sort items had proved to be controversial and which had not and appropriate debates could be encouraged. We conducted 5 groups in all between April and October of 1993 at various towns and cities in the South of England: one among employers, three groups among different subgroups of the general population, and one among researchers who were part of the project's ED Steering Group. Individuals from the general population and employers represented a carefully balanced quota sample selected by field interviewers through door-to-door screening. A handful of other researchers from the ED were also asked to complete the card sort task. This brought our total number of card sort respondents to 49.

#### 4. Substantive and Methodological Findings of the Survey Approach

**Card Sort Data.** Cognitive laboratory data, due to their wealth of information and small sample sizes, typically lend themselves to qualitative rather than quantitative analysis. With our sample of 49 respondents, we employed a combination of both analysis methods.

As a first step to learning how people viewed the concept of training, we examined the extent to which each item was classified as *training*. Table 1 orders the 30 items into 7 levels based on the amount of support it received. This ordering provided information with which to examine the validity of the ED's assumptions. For example, the pattern of results clearly supports Assumption 2. The items forming Level 1 all represent formal courses and are much more likely to be classified as *training* than the informal activities items (C1-C6), the apprenticeship items (C7-C11), and the job experience items (C12-C15). It is also interesting to note that only C24 of the Level 1 group clearly points out that the *training* will result in a qualification. This suggests that perhaps earning a qualification may not be a necessary condition for Assumption 2.

These data also lend support to Assumption 5. This is shown by the fact that self-initiated/funded

activities from Level 3 are less likely to be classified as *training* than items such as those contained in Level 1. It should be noted, however, that these

Table 1 Percentage of R's Identifying Card Sort Items as *Training*, by Individual Items

Item	Description
<b>Level 1: 90-98 per cent classify as training</b>	
C16	A 1 day course to learn the photo copier
C17	A 1 day course to learn how to use a chain saw safely
C24	Course for qualification in accountancy
C27	Day release for typing proficiency
<b>Level 2: 80-89 per cent classify as training</b>	
C7	Working in each department, so as to learn how the company works
C23	Studying to be a medical doctor
C25	Studying to be a clinical psychologist
<b>Level 3: 67-78 per cent classify as training</b>	
C11	Being advised how to handle new order
C20	Self-initiated/funded vocational qualification
C26	Youth Training Scheme (YTS)
C28	Self-initiated/funded course peripherally related to work
C30	Self-initiated/funded, work connection no qualification
<b>Level 4: 37-53 per cent classify as training</b>	
C1	Talking with co-workers
C2	Receiving help from a more experienced worker
C3	Asking a more experienced worker to check work
C4	Reading journal articles
C5	Reading manuals
C9	Watching and helping as first step of an apprenticeship
C15	Being given short tasks with explanation before each
<b>Level 5: 26-37 per cent classify as training</b>	
C19	Self-initiated/funded GCSE study (similar to high school equivalency exam)
C21	Self-initiated/funded art history, after retirement
C22	Self-initiated/funded foreign language study for enjoyment
C14	Having grammar and spelling corrected
<b>Level 6: 11-24 per cent classify as training</b>	
C6	School of life
C8	Menial tasks as first step of an apprenticeship
C10	Basic instructions (what to wear, how paid) as first step of new job
C12	Gaining experience actually doing the job
C13	Learning by mistakes
C29	Self-taught mechanic
<b>Level 7: 2 per cent classify as training</b>	
C18	Keep-fit class

The text of this table represents descriptions of the content of the card sort items, not the actual wording of the items.

three Level 3 items were still endorsed by a majority of people as *training*. In contrast, what appears to be a key discriminator is whether or not the activity is vocationally linked. For example, C20, C28, and

C30 (from Level 3) clearly take precedence over C19, C21, and C22 (from Level 5) which do not have a clear vocational link. It is also the lack of a vocational link that disqualifies C18 (from Level 7) from being *training*.

The fact that the education items do not fall either around the 98 per cent end or the 2 per cent end of the table suggests that there is ambiguity between the concepts of education and *training*. This supports Assumption 6.

Exceedingly similar results were found with the analysis of the 12 vignette examples which had been commissioned on the random Omnibus survey, each among a sub-sample of 500 respondents. The linguistic findings also supported the same ranking. The similarity of findings from these different sources of data lends strength to the conclusions.

---

**Table 2** Summary of the Likelihood of Each Group Endorsing Card Sort Items as *Training*

---

For self-initiated/funded and educational activities with a vocational link:

Employer < General Population < Researchers

For on-the-job items:

Employer > General Population

For experiential items and education for pleasure:

General Population = Researcher  
(low likelihood of endorsement)

For classic *training* items:

Employers = General Population = Researchers  
(high likelihood of endorsement)

For informal activities:

ED Steering Group researchers > Other ED researchers

For informal activities, apprenticeship tasks, job experience, and some education items:

Large company employers < Other employers

---

The card sort data were also used to investigate Assumptions 1, 3, and 4. Summing across all items and all group members suggested that researchers had a broader definition of *training* than either the general population or employer groups at the  $p < .05$  level, supporting Assumption 1. In contrast, employers appeared to have a narrower definition of *training* than the general population, contradicting

Assumption 3. Examination of the tables for the individual items suggested the patterns described in Table 2. Variation among members of the general population was also found with respect to their age, social grade, and the industry of their employment.

**Other Data Sources.** Due to page limitations of this document, the results on the omnibus data will not be discussed here. These data, however, allowed Assumption 4 to be explored more fully. The results expanded upon and confirmed findings from the card sort data. The qualitative data collected during the focus group discussions were also very useful. Some of the qualitative findings are listed in the summary in Section 7.

### 5. Some Findings from the Linguistic Approach

The linguistic approach first pointed to the distinction between the verb *to train* in its various forms, and the noun *training*. From the corpus linguistic work it appeared that the application of the noun *training* is wider than that of the verb. The verb is often used in the context of specific occupations, less often about specific skills. The context for the noun, however, is often vocational and often about general skills.

The results also suggested that when used in questionnaires, particular grammatical patterns may be expected to elicit particular sorts of responses. For example, 'What are you trained as?' will elicit an occupation name, with occupation as a result of formal training period. Respondents whose occupation does not fit their idea of *trained as . . .* might reply 'nothing'.

The linguistic approach also examined how people talk about *training* when they don't use the word *training* (or its related word forms). Some conclusions from this strand were that:

- there are a large number of expressions which respondents do indeed use to talk about their training, education and qualifications; and
- many of these expressions involve high frequency, all-purpose verbs (such as *do, go, get*, e.g. 'do an apprenticeship', 'go on a course', 'get a diploma'.)

### 6. Revised assumptions

In general, Assumptions 1, 5, and 6 were clearly supported by both the linguistic and survey data. Assumption 2 was supported by the linguistic data and slightly qualified by the survey data. Whether or not an activity leads to a qualification is less relevant than whether or not it is vocationally linked. Assumption 4 was partially supported by the survey data and generally not supported by the linguistic data. It should be noted that it will be important to explore individual experiential factors that affect

people's definitions. Assumption 3 was not supported by the survey data. The situation surrounding employers views is complex and deserves further research.

## 7. Implications for Survey Design

Development of Training Questions. Data from the card sorts, vignettes and focus groups discussions confirm that the general adult population does not interpret the word *training* in a way which is sufficiently broad and consistent for it to be used as a key element in standard questions about vocational learning and *training*. Asking respondents direct questions about their training experiences during a particular reference period (as is done in some surveys) can lead to substantial under-reporting of relevant events.

A similar conclusion was reached by the linguistic approach: the word *training* and other forms of the verb *train* are not the best words to use in questions about training. The linguistic approach found that no uses of *train* seem to encompass the idea of 'on-the-job, learning by experience, informal training'. All forms of questions which use *train* or *training* therefore appear to restrict respondents' replies and lead to under-reporting. One questionnaire design solution would be to ask standard direct questions about formal courses, but then to prompt an extended list of key informal activities. This approach ensures that such activities are not missed.

Conclusions from the linguistic approach, however, prompted a more radical divergence from past models by developing a line of questioning which avoids the use of the word *training*. Choosing from among other commonly used expressions, questions involving the word *learn* appear to workable. Typical uses of *learn* appear to prompt respondents to consider and report on the widest range of training-like activities relevant to their own working life. At the same time, while it is a common word, *learn* does not have the possible sociolinguistic problems associated with the high frequency verbs like *do* and *get* (these problems are described in Campanelli et al 1991). A possible formulation is for example: 'How did you learn to do your job?' Another favoured formulation could be to use the word *acquire*. It also elicits general responses, and among its typical contexts are work related items such as *skill*, *competence* and *knowledge*. A possible formulation is: 'How did you acquire the skills you need for you work?' These two questions were used as the basis for a specific pretest.

Questionnaire Pretest. One goal of the project was to develop a training question module for the new UK Working Lives Survey (WLS), co-sponsored by the ED and several other UK government departments. A small-scale cognitive laboratory type pretest was held in June with a carefully balanced quota sample of 16 respondents from a city west of London. As with the focus groups, respondents were recruited by SCPR interviewers as part of their standard fieldwork duties. The two project researchers (who were also experienced survey interviewers) travelled to the homes of the recruited respondents and administered the pretest questionnaire followed by a cognitive-style debriefing session. Respondents were alerted in advance to the two parts to the interview.

From the pretest, it seemed reasonably clear that the 'learn to' and 'acquire' lines of approach are good ones which avoid some of the problems which arise where the word *training* is used. The use of a showcard to prompt for potentially under-reported activities was also critically important.

## 8. Summary and Conclusions for Survey Design

We conclude with some general recommendations to survey designers which are based on the research.

- 1 Avoid the use of the word *training* (and its associated word forms). These are not the best words to use in questions about training.
- 2 An approach which asks 'How did you learn to do your job?' or 'How did you acquire the skills you need for your work?' is a good one which avoids some of the problems which arise where the word *training* is used.
- 3 Avoid trying to re-educate respondents to think in terms of researchers' definitions of training. It is likely this will be complex and only create confusion rather than clarity. Instead use a showcard to prompt for potentially under-reported training activities. Use of a showcard is critically important.
- 4 While researchers or training professionals distinguish between the process and the product of training, many of the ways people talk about training can involve both of these elements. Be clear about what is wanted in terms of process versus product. Is the survey more interested in measuring length of training or number of skills and competencies?
- 5 Respondents from lower socio-economic occupations may need particular encouragement, as they may not realize at first that their jobs involve skills which they had to acquire.
- 6 Be aware that respondents' individual

classification of the cognitive category 'training' will be variable along a number of socio-demographic factors. There are also large differences between the views of employers, researchers, and members of the general population with respect to training.

7 Be aware that there is a built-in vagueness about the word *training*. It can operate at 3 distinct levels of meaning:

- training in a general sense, with no specific job or skill in mind
- training for a specific occupation
- training for a specific skill needed for a specific occupation

The topic or sponsorship of a particular survey may indirectly (and not always helpfully) suggest to respondents which of these 3 levels of meaning is being referred to.

8 Be aware that although it is perfectly understandable, there is no generally-used expression *self-trained*. For some people training is seen as something that is provided for (or even imposed upon) the individual employee, rather than something which he or she undertakes on personal initiative.

9 Be aware that many respondents feel it is more important that there is deliberate and systematic teaching in order for an activity to be considered *training*, than for the activity to lead to a formal qualification. Similarly 'hands-on' practice of skills, under supervision, was seen as a prerequisite for 'proper' training, not just watching. *Informal learning experiences*, such as more experienced workers giving help and advice to a learner could lead to the perpetuation of bad practice and were therefore generally not seen as *training*.

10 There is a fuzzy boundary between education and training. Collect data on both to ensure that everything has been covered.

## REFERENCES

Belson, W.A. (1981). *The Design and Understanding of Survey Questions*. London: Gower.

Campanelli, P.C., E.A. Martin, and J.M. Rothgeb. (1991). The Use of Respondent and Interviewer Debriefing Studies as a Way to Study Response Error in Surveys. *The Statistician*, 40, 253-264.

Campanelli, P., J. Channell, L. McAulay, A. Renouf, and R. Thomas. (1994). *Training: An Exploration of the Word and the Concept with an Analysis of the Implications for Survey Design*. Sheffield: UK Employment

## Department.

Drew, P. and Heritage, J. (eds.) (1992). *Talk at Work: Interaction in Institutional Settings*. Cambridge: Cambridge University Press.

Forsyth, B. and J. Lessler. (1991). *Cognitive Laboratory Methods: A Taxonomy*. In *Measurement Errors in Surveys*, eds. P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, N.A., and S. Sudman. New York: John Wiley and Sons, Inc.

Martin, E.A., R. Groves, J. Mattlin, and C. Miller. (1986). *Report on the Development of Alternative Screening Procedures for the National Crime Survey*. Washington, DC: Bureau of Social Science Research.

National Center for Health Statistics. (1989a). *Questionnaire Design in the Cognitive Research Laboratory, Series 6: Cognition and Survey Measurement, No. 1*, DHHS Publication No. (PHS) 89-1076. Hyattsville, MD: US Department of Health and Human Services.

National Center for Health Statistics. (1989b). *Social Cognition Approach to Reporting Chronic Conditions in Health Surveys, Series 6: Cognition and Survey Measurement, No. 3*, DHHS Publication No. (PHS) 89-1078. Hyattsville, MD: US Department of Health and Human Services.

Neisser, U. (1987). *Concepts and Conceptual Development*. Cambridge: Cambridge University Press.

Robson, J. (1992). *Working Lives and Training*. University of Sheffield, UK: Qualitative and Quantitative Studies in Education Research Group.

Rosch, E. (1975). Cognitive Representations of Semantic Categories. *Journal of Experimental Psychology*, 104, 192-233.

Rosch, E. and B. Lloyd (eds.) (1978). *Cognition and Categorization*. New York: Lawrence Erlbaum.

Rossi, P.H. and A.B. Anderson. (1982). *The Factorial Survey Approach: An Introduction*. In *Measuring Social Judgments: The Factorial Survey Approach*, eds. P. Rossi and S. Nock. Beverly Hills: Sage Publications, Inc.

Sinclair. (1991). *Corpus Concordance Collocation*. Oxford University Press.

1 This work was funded by the UK Employment Department. We would also like to thank several persons at both SCPR and the Research Development Unit for English Studies (then at the University of Birmingham) for their helpful contributions to the project.