Proceedings of Statistics Canada Symposium 2001
Achieving Data Quality in a Statistical Agency: A Methodological Perspective

TESTING THE MODE EFFECTS IN THE FINNISH CONSUMER SURVEY

Pertti Kangassalo and Markku Heiskanen[1]

ABSTRACT

From January 2000, the data collection method of the Finnish Consumer Survey was changed from a Labour Force Survey panel design mode to an independent survey. All the interviews are now carried out centrally from Statistics Finland's CATI Centre. There have been suggestions about the survey mode influencing the respondents' answers. The aim of this paper is to analyse the extent of obvious changes in the results of the Consumer Survey. This is accomplished with the help of a pilot survey. Furthermore, we discuss the interviewer's role in the process. The analysis is based on cross-tabulations, chi-square tests and multinomial logit models. We will show that the new survey method produces somewhat more optimistic estimations and expectations concerning economic matters than the old one did.

KEY WORDS: Mode effects; CATI; Interviewers; Consumer Survey

1. INTRODUCTION

There are earlier studies showing that the survey mode has a central role in influencing the respondents' answers. In particular, there can be great differences between a postal survey and a telephone interview or between a telephone survey and a face-to-face field interview (de Leeuw, 1993). The topic of the survey is also relevant when the mode effect is analysed. It can be assumed that the data collection mode effect is quite minor if the questions mainly concentrate on factual issues, but the difference can be greater in respect of questions handling more sensitive issues (Nicholls et al., 1997; Heiskanen and Ahlqvist, 1997).

Normally the mode effect is used to describe differences in coverage, nonresponse or measurement error between different data collection methods (Nicholls et al., 1997). It is also worthwhile to note that «it is important to make a distinction between the method or means of data collection (e.g. interview, self-administered questionnaire) and the technology used to capture the data» (de Leeuw and Collins, 1997).
Statistics Finland introduced the Consumer Survey (CS) in November 1987. From October 1995, the data have been collected monthly in accordance with the harmonised EU data collection method (European Commission, 1997). The main function of the CS is to study the consumers' qualitative valuations and expectations concerning their own and the general economic situation, i.e. the economic sentiment.

Until December 1999, the data for the CS were collected together with the data of the Labour Force Survey (LFS). Six rotating panels were then used in the CS. The same person was asked the same questions three separate times at six-month intervals, and each month a third of the target persons were new. Statistics Finland's field interviewers conducted the telephone interviews (CATI) around the country.

As of January 2000, the CS has been based on a totally new sample each month. The same sample is used for the data collection of the Finnish Travel Survey, and some occasional surveys as well, but the CS questions are always asked first. All the interviews are now conducted centrally from Statistics Finland's CATI Centre in Helsinki (Statistics Finland, 2000-2001).

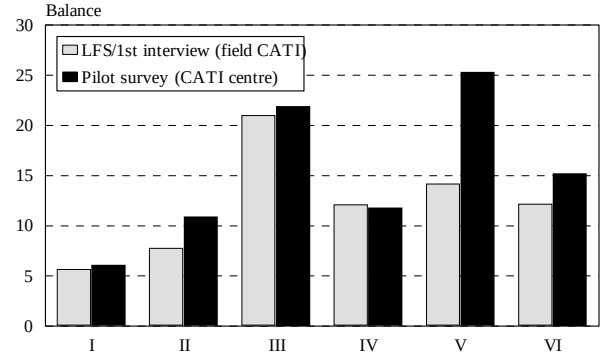
The aim of this paper is to analyse whether the altered data collection method has influenced the responses of the Finnish CS. The statistical significance of the mode effect is examined. This is accomplished with the help of a pilot survey that was carried out at the same time as the LFS-based survey in November 1999. So

[1] Statistics Finland, P.O. Box 6A, FIN-00022 Statistics Finland, Finland; pertti.kangassalo@stat.fi.

as to avoid the possible impact of the panel design, only the data of the first rotation group, the first-timers, of the LFS-based survey are used in comparisons.
In this kind of case, one might rather suppose that there would not be any significant differences in the survey results. First, the same structural questionnaire is used in both methods and the measurement is as standardised as possible, i.e. asking the questions is more or less routine for the interviewers. Second, the statistical quality demands are the same for both survey types. Third, before starting the interviews the interviewers of the CATI Centre received training in the interviewing technique and the contents of the CS.

However, basing on cursory comparisons of results between successive months and the pilot survey around the turn of the year 2000, the responses of the CS seem to have changed. More precisely, the respondents' evaluations concerning their own households' economic situation and the favourability of the time to purchase consumer durables and to save, in particular, may have become more positive, even though the actual questions of the CS have remained exactly the same[2]. As Figure 1 shows, this has had an effect on the most important outcome of the survey, i.e. the Consumer Confidence Indicator[3].

Figure 1. CCI and its components surveyed with two different methods in the Finnish Consumer Survey



There are various possible reasons why these changes in observed consumer attitudes might have occurred. Of course, the first possibility is that the attitudes have really turned to a more positive direction. However, on grounds of the findings from the pilot survey this is unlikely to fully explain the changes. The second possible explanation is the change in the mode of the survey. It is clear that in this case there is more than «one mode effect» that can explain the changes. For example, differences in nonresponse between the data collection methods could be one reason. However, even though the nonresponse of the CS increased to some degree (except for the pilot survey) and the relative share of noncontacts clearly grew as the method was altered, the nonresponse by different background factors remained the same (Kangassalo and Notkola, 2000). Therefore, nonresponse can be counted out.

2. DATA AND METHOD

We used two data sets for the analysis in this paper (Table 1). The first one comes from the CS that was carried out together with the LFS in November 1999. Statistics Finland's field interviewers (n=147) in the provinces conducted the telephone interviews. The LFS panel design was used, but for this study only the first-timers, i.e. 598 respondents, were picked out to allow best possible comparisons with the other data set, i.e. the data of the pilot survey that was based on an independent random sample of individuals. In the pilot survey the interviewers of Statistics Finland's CATI Centre (n=13) conducted all the interviews by tele-

[2] Quite similar experiences have also been gained from the Consumer Survey in Sweden (Statistics Sweden, 2000).
[3] The CCI is an arithmetic average of the balance figures of five questions. Balance figures are typical for qualitative surveys; they are obtained by deducting the weighted proportion of negative answers from that of positive answers.

phone. The 582 observations in the pilot survey data set represent exactly the same time period (week 47, Nov. 1999) as the LFS.

For both surveys the target area was the whole Finland and the respondents represented the country's 15 to 74-year-old population by gender, age, province and native language. Systematic random sampling was used to extract the sample from the updated Central Population Register. The sorting system of the sampling frame was based on geographical population density.[4]

The results of the surveys were afterwards weighted against the total population by means of sample weights in exactly the same way for both surveys. The weights were established by using the inclusion probability of each observation in the sample and finally by a calibration method. For individual weights the marginal variables were the respondent's gender, age group and major area. Correspondingly, in household weighting, number of household members by age group and major area served as the marginal variables.

The wording of the questions was exactly the same for both surveys (EU harmonised questionnaire) and the same answer alternatives were given to the respondents, i.e. (in most cases) two positive (much better/a little better, a lot more/a little more, very good/quite good, yes/maybe), two negative (a little worse/much worse, a little fewer/a lot fewer, quite bad/very bad, maybe not/no), neutral (same, neither) and «Don't know» categories.

Table 1. Analysis data sets

Old method	**New method**
- LFS 11/1999, week 47	- Pilot survey 11/1999, week 47
- Field CATI, 147 interviewers	- CATI Centre, 13 interviewers
- First-timers of the panel (1st interview)	- Independent sample («first-timers»)
- n = 598	- n = 582
- Nonresponse rate = 27%	- Nonresponse rate = 23%
- Average interviewing time = 11 min	- Average interviewing time = 15 min
- EU harmonised questionnaire	- EU harmonised questionnaire
- Results weighted (calibration method)	- Results weighted (calibration method)

The analysis of the differences between the survey results below is based on normal cross-tabulations, chi-square tests and multinomial logit models (e.g. Everitt and Dunn, 1983). The logistic analysis was carried out by means of the SAS/GENMOD procedure. As explanatory factors we used the following demographic and socio-economic variables (see frequencies in Appendix) which could be linked to the data sets from administrative registers:

- Gender
- Age: 1. 15-34, 2. 35-54, 3. 55-74 years
- Education: 1. basic, 2. secondary, 3. higher education
- Household income: 1. lowest quartile, 2. second lowest/second highest quartile, 3. highest quartile
- Employment: 1. employed, 2. unemployed, 3. other
- Region: 1. Greater Helsinki, 2. the rest of Finland.

3. RESULTS

3.1 Chi-square tests for differences in answer distributions

From among the many questions included in the CS questionnaire we chose the 17 most essential ones for the analysis. These questions concern both the general economic situation in Finland (5 questions) and the

[4] The sampling method can be compared to the use of simple random sampling with the addition that the sample is geographically self-weighting.

respondent's own economic situation (4) as well as the favourability of the time to purchase consumer durables, save and raise a loan (3). Questions about households' intentions to buy a dwelling or a car, repair a dwelling, save, and raise a loan were also included in the study (5).

According to the chi-square tests of the answer distributions, clearly in more than half of the questions, i.e. in 11 variables out of 17, there appeared to be a statistically significant difference at 95% level (Table 2). The share of differing cases was the highest for questions concerning the favourability of the time (3 out of 3 questions) and the respondent's own economy (3/4). Correspondingly, this share was the lowest for questions on the household's diverse intentions (2/5).

When calculated separately for male and female respondents the results were quite different compared to the common figures (mostly for reasons connected with the small sample sizes), but very similar with one another. Still, when considering the different questions we found some variation between the genders. With both genders, questions concerning the favourability of the time to purchase consumer durables, intentions of spending money and doing basic repairs to a dwelling were among the delicate ones. With men there also appeared to be a significant effect in questions about saving possibilities, the financial situation of the household, and about the purchase of a car. Correspondingly, female respondents seemed to have answered differently depending on the mode to questions concerning consumer prices now, the development of unemployment, favourability of the time to save, and intentions to raise a loan.

It is also interesting to analyse the directions of the changes in the results. Balance figures (see above, Chapter 1) were used for determining the signs. Table 2 shows that a clear majority of the changes were positive ones, in particular those on the significant side. Therefore, basing on the chi-square tests we can say that the change of the data collection method has had a statistically significant effect on the CS results and in most cases this effect has been positive, i.e. with the new method the respondents' opinions on economic matters have grown somewhat more optimistic.

Table 2. Number of questions with significant/insignificant differences between the old and new survey method according to chi-square tests

Significant differences ($p<0.05$)	Positive change	Negative change	**Total**
All	8	3	11
Males	4	2	6
Females	4	3	7
Insignificant differences			
All	3	3	6
Males	6	5	11
Females	7	3	10
Total	11	6	17

3.2 Multinomial logistic regression analysis

As the second step, we carried out a multinomial logistic regression analysis by means of generalised linear modelling. We restricted the analysis to the variables which had experienced the biggest (positive) changes from the new data collection method. These were the respondent's own economic situation in twelve months' time, favourability of the time to purchase consumer durables and save, and intentions to spend money on basic repairs to dwelling in the next twelve months. Above, the chi-square tests were conducted with full response categories, but for the following analysis the answer distributions were treated in a way that the two positive (much better/a little better, very good/quite good, yes/maybe) and two negative (a little worse/much worse, quite bad/very bad, maybe not/no) categories, respectively, were combined together to ensure sufficient observations for reliable results. Moreover, the «Don't know» answers were excluded from

the analysis. Although the share of these answers is clearly lower for the centralised survey, there proved to be no variation between the methods by the explanatory variables.

Respondent's own economic situation

Basing on the chosen model there was no statistically significant difference between the survey methods in general, when the effects of all other factors were standardised. Nevertheless, we found a statistically significant interaction effect ($p=0.0361$) by the level of education: interviewees having only basic education (or their education was unknown) were more optimistic in the pilot survey than in the LFS-based survey (Table 3). At this lowest educational level, 21.5 per cent of the respondents had a positive view concerning their own economic situation when asked with the new method, but only 12.1 per cent thought so in the LFS-based survey. Especially male respondents proved to be very sensitive to this education effect since their corresponding shares at the lowest educational level were 25.6 and 10.3 per cent.

We also found some evidence of differences between the methods by the respondent's age ($p=0.0676$), i.e. in the oldest age group (55-74 years) the optimistic shares were 13.5 and 6.1 per cent for the pilot survey and the LFS-based survey, respectively. Again, a clear difference (16.9 vs. 1.8) was found among the males of this age group.

For both methods respondents representing different gender or age groups gave divergent answers about their own economy ($p=0.0478$ and <0.0001 resp.).

Favourability of the time to purchase consumer durables

There was no variation between the survey methods in the case of this variable and all the interaction effects remained insignificant. However, for both methods the answer distributions differed significantly ($p=0.0163$) from each other by gross income of respondent's household.

Favourability of the time to save

There appeared to be a significant difference ($p=0.0125$) between the two methods so that the new method produced more favourable views. The shares of positive opinions were 57.6 and 50.3 per cent, respectively. As with the case of respondent's own economic situation above, we found that the answer distributions of the two surveys differed by respondent's educational level ($p=0.0297$). Again, the phenomenon was the clearest at the lowest educational level: 50.1 versus 35.4 per cent of the respondents had positive views about saving. Furthermore, with male respondents with basic education only the corresponding figures were 56.5 and 29.4 per cent. For this response variable, the opinions at the secondary education level were also somewhat more positive when studied with the centralised mode.

With both methods the answers by respondent's age were different ($p<0.0001$), i.e. younger interviewees thought more positively about saving.

Intention to spend money on basic repairs to dwelling

There was significant difference between the survey methods in general ($p=0.0282$), i.e. with the new method more intentions (57.6 vs. 50.3%, males: 60.9 vs. 51.8%) came out, but no factor/method interaction effects appeared. Nevertheless, with both methods the respondents' intentions varied significantly by region and household income ($p=0.0040$ and <0.0001 resp.).

Table 3. P-values for differences between the survey methods by different background factors, multinomial logit model

	Own economic situation in 12 months' time (better/same/worse)	Favourability of time to purchase consumer durables (good/neither/bad)	Favourability of time to save (good/bad)	Intention to spend money on basic repairs to dwelling in the next 12 months (yes/no)
Gender	0.6841	0.5409	0.9945	0.2548
Age	*0.0676*	0.3495	0.8852	0.5100
Education	*0.0361*	0.3180	*0.0297*	0.9084
Household income	0.7736	0.6746	0.2403	0.3594
Employment	0.6606	0.4356	0.3804	0.3451
Region	0.5592	0.4646	0.4800	0.2356
Survey mode	0.1942	0.2386	*0.0125*	*0.0282*
Answer distributions, % («Don't know» answers excluded)				
LFS/1st interview	25.4/62.9/11.7	45.7/25.2/29.1	50.3/49.7	13.4/86.6
Pilot survey	28.3/63.7/8.0	51.6/23.5/24.9	57.6/42.4	18.7/81.3
N	1,148	1,072	1,117	1,155

As mentioned above, we were forced to combine some answer categories for the regression analysis. More of significant differences might have emerged had we had the possibility to use larger samples for the study. This is indicated by the fact that some displacements had also occurred inside the positive side (e.g. much better/a little better) and the negative side (e.g. a little worse/much worse), respectively, not only between the two sides.[5]

4. CONCLUSIONS

We found that there were statistically significant differences in the answer distributions of two-thirds of the CS's most essential questions between the two data collection methods of the Finnish CS, i.e. the centralised CATI and the LFS-based field CATI. Furthermore, in three out of four significant cases the sign of the change was positive. Therefore, in general, we can conclude that the changeover to the centralised CATI method in the Finnish CS has turned the survey responses toward a more optimistic direction.

We also found that depending on the response variable the answer distributions varied by different background factors. We were particularly interested in possible interaction effects between the background factors and the survey methods and, indeed, we discovered some evidence of these: the respondents with only basic education, and of them especially men, gave more optimistic answers to certain questions with the new survey method. This finding is known from some other Finnish survey studies as well (e.g. Ahola et al., 1995; Heiskanen and Ahlqvist, 1997), and it could be explained with the so-called social desirableness that respondents with low educational levels seem to be particularly sensitive to. Moreover, it could even be a question of some kind of «Finnishness» or «fear of administration», as Hofstede (1980) describes it, which may become more apparent when a telephone interview is conducted in a centralised mode. The fact that the interviewers are located in the capital may increase the distance and formality between the respondent and the interviewer.

The local dialect (or intonation) the field interviewers often use in provinces or the absence of it in Helsinki-centred telephone interviewing could be linked to the way respondents express themselves. Furthermore, there are differences in the interviewers' average age[6] and experience between the two modes. According to

[5] Knowing the restrictions of the sample sizes available, we still carried out a cursory analysis with the full response categories and did not find any additional evidence of differences.

[6] The average age of the field interviewers is about 50 years but that of the CATI Centre interviewers only half of it.

Houtkoop-Steenstra and Antaki (1997), if respondents give unclear or inconsistent answers interviewers may select the most optimistic answer unless there are clear indications from the respondent that another answer is more in accord with what the respondent means. This could well be typical for new and inexperienced interviewers who may try to please the respondent.

On the other hand, the independence of the centralised CS, or the current association with the Travel Survey, instead of the «serious-type» LFS that concentrates on employment and unemployment, might have caused some effects similar to the above.[7] Furthermore, in the new method the CS questions are always asked first of the two surveys, whereas in the past they never preceded the questions of the LFS. However, the questionnaire still starts with some background questions concerning, for example, the main activity (employment) of the respondent and the income class of the respondent's household which brings the new survey close to the old one. Still, the atmosphere of the survey may now be somewhat more open when compared to the LFS method.

Could some other things possibly explain the differences in the survey results? As such, the change from the field to the CATI Centre may not have caused any differences because in both methods the same basic technique is used; i.e. interviewers, telephones and computer-assisted recording. As mentioned above in Chapter 1, we can at least conclude that the answer is not the different response rates of the two survey methods.

Finally, which one of the two methods actually produces more accurate results? On grounds of the discussion above, the results seem to be more than enough optimistic with the CATI Centre method particularly in the case of low educated men. For men - more often than for women - getting along in economic sense is an important value. Autumn 1999 represents the time of economic boom in Finland, which emphasizes the expectations to be successful. During a recession the social desirableness might influence in an opposite way, i.e. lower social groups might stress their (economic) problems.

After all, in a researcher's point of view managing the interviewing process and controlling its quality is easier in the CATI Centre than it was in the field mode. This probably has positive effects on the survey in the long run.

REFERENCES

Ahola A., Djerf K., Heiskanen M. and Vikki K. (1995), Elinolotutkimus 1994 - Aineiston keruu, Statistics Finland - Memorandum 1995:2, Helsinki. (In Finnish)

De Leeuw E. (1993), «Mode Effects in Survey Research, A Comparison of Mail, Telephone and Face to Face Surveys», *BMS*, 41, pp. 3-15.

De Leeuw E. and Collins M. (1997), «Data Collection Methods and Survey Quality: An Overview», in L. Lyberg et al. (eds.) *Survey Measurement and Process Quality*, New York: Wiley.

Djerf K. and Takala K. (1997), Macroeconomy and Consumer Sentiment: Performance of the Finnish Consumer Barometer after Ten Years, Bank of Finland Discussion Papers 20/1997, Helsinki. Also published in Oppenländer K.H., Poser G. and Weller S. (eds.) *Selected Papers Submitted to the 23rd CIRET Conference in Helsinki*, Aldershot: Ashgate.

European Commission (1997), European Economy, The Joint Harmonised EU Programme of Business and Consumer Surveys, Reports and Studies No 6, Brussels.

Everitt B.S. and Dunn G. (1983), *Advanced Methods of Data Exploration and Modelling*, London: Heinemann Educational Books.

Groves R.M. (1989), *Survey Errors and Survey Costs*, New York: Wiley.

[7] In literature these effects are known as «context effects» (see e.g. Groves, 1989, pp. 478-481).

Groves R.M. and Couper M.P. (1998), *Nonresponse in Household Interview Surveys*, New York: Wiley.
Heiskanen M. and Ahlqvist K. (1997), «Experiences of Two CAPI-PAPI Comparisons in Finland», paper presented at the 4th International BLAISE Users' Conference, Paris, France.

Hofstede G. (1980), *Culture's Consequences: International Differences in Work-Related Values*, Beverly Hills, CA: Sage Publications.

Houtkoop-Steenstra H. and Antaki Ch. (1997), «Creating Happy People by Asking Yes-No Questions», *Research on Language and Social Interaction*, 30(4), pp. 285-313.

Kangassalo P. and Notkola V. (2000), «Centralised CATI Method in the Finnish Consumer Survey - Testing the Effects of Nonresponse», paper presented at the 25th CIRET Conference, Paris, France.

Laaksonen S. (ed.) (1996), International Perspectives on Nonresponse - Proceedings of the Sixth International Workshop on Household Survey Nonresponse, Statistics Finland - Research Reports 219, Helsinki.

Nicholls W.L. II, Baker R.P. and Martin J. (1997), «The Effect of New Data Collection Technologies on Survey Data Quality», in L. Lyberg et al. (eds.) *Survey Measurement and Process Quality*, New York: Wiley.

Statistics Finland (2000-2001), Consumer Survey publications and working papers. (Mostly in Finnish)
Statistics Sweden (2000), Hushållens inköpsplaner, januari 2000, Pressmeddelande från SCB Nr 2000:021. (In Swedish)

Appendix. Frequencies by different background factors and by data set

Background factor	LFS/1st interview	Pilot survey
Gender		
Male	286 *47.8*	294 *50.5*
Female	312 *52.2*	288 *49.5*
Age		
15-34 years	211 *35.3*	199 *34.2*
35-54	245 *41.0*	220 *37.8*
55-74	142 *23.7*	163 *28.0*
Education		
Basic	205 *34.3*	187 *32.1*
Secondary	109 *18.2*	120 *20.6*
Higher	152 *25.4*	162 *27.8*
Unknown	132 *22.1*	113 *19.4*
Household income		
Lowest quartile	128 *21.4*	121 *20.8*
Second lowest or second highest quartile	256 *42.8*	278 *47.8*
Highest quartile	110 *18.4*	144 *24.7*
Unknown	104 *17.4*	39 *6.7*
Employment		
Employed	352 *58.9*	311 *53.4*
Unemployed	39 *6.5*	35 *6.0*
Other	207 *34.6*	236 *40.6*
Region		
Greater Helsinki	103 *17.2*	87 *15.0*
The rest of Finland	495 *82.8*	495 *85.0*
Total	598 *100.0*	582 *100.0*