A study of Systematic Errors in Survey Research: The effect of the perception of other people's opinions

C. Bruinsma
W.E. Saris
I.N. Gallhofer

Abstract.

Interviewer effects on survey responses can be interpreted as effects which occur as the result of respondent's perception of opinions by other people.
However, this study shows that the effects of such perception variables are not significantly different from zero.
This finding is far from incidental but throws great doubt on existing evidence in the literature about interviewer effects.
Other systematic errors, however surely exist, but considering the insignificant effects of perception variables their causes remain unclear.
This study also shows that categorical evaluations can not be repeated without strong memory effects, alternative procedures such as magnitude estimation are less sensitive for such errors and should be used more widely.

Response errors in survey research have been studied frequently and these studies (e.g. Johnson, Dijkstra and Elsinga, Sudman and Bradburn) have been amply reviewed under the following headings:
a. the effect of social desirability (e.g. Edwards, Crowne and Marlowe, Phillips, 1971, 1973); b. the effect of interviewer characteristics (e.g. Weiss, Henson et.al.); c. the effects of interviewer opinions (e.g. Hyman et.al., Freeman and Butler).
Since only significant findings were published we do not know as yet whether such response errors could arize by chance alone. In addition, it remains unclear why their effect occur at all: the processes by which the errors are caused remain largely unspecified or untested. Hetebrij suggested that the effects can be interpreted by the variable "perception of the opinion of other persons". The respondent gets indications of this perception from characteristics of the other person. Age, sex, status, clothes might be used as clues for example. The other person can be the interviewer but also the family, respondent's collegues of the group in which he or she works, the management, etc.. These other persons can also transmit their own opinion in some way or another to the respondent.
In fact it is even not necessary that the other persons are present in the interview situation as long as the perception of their opinions are salient for the respondent (Hyman, et.al.).
We do not expect the effects of the characteristics of the other persons on the perception of the respondent to be very strong. The effect of the perception variable on the responses might be much stronger. But if one of the two is close to zero the indirect effect of the characteristics of the other persons on the responses will be zero too.
Given the hypothesized proces the perception of the "opinion of relevant others" would be a more important variable in this proces.
If the effect of this variable on the answers of the respondent is very weak we do not expect that the earlier mentioned variables are of much importance.
On the other hand a strong effect of this perception variable does not indicate effects of the other variables as the perception can result from completely different variables.
Therefore we chose to study the effect of a perception variable on the answers of the respondent in survey situations.
The topic chosen to test this hypothesis is the evaluation of job perfor-

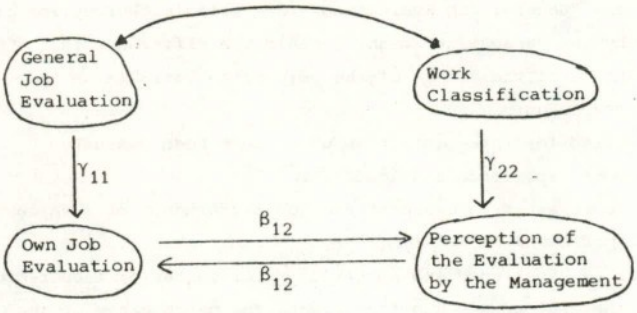

Diagram 1. A path diagram of the model of this study.

mance by workers. It is attractive as one can explicitly ask respondents to evaluate their own qualities as much as one can ask the evaluation of a large variety of qualities of which some are their own qualities. In this way the respondents give twice an evaluation of the same qualities.

The first evaluation is called "Own Job Evaluation", the second is called "General Job Evaluation". Normally one expects that evaluations of the same characteristics lead to the same scores except for random measurement error. However if the respondent also knows in one case that his own qualities are evaluated he might respond differently in order to obtain as much advantage as possible from these evaluations. We thought that in that case the "Perception of the Evaluation of his Qualities by the Management" could play an important role. External information with respect to the evaluation by the management was available to the respondent: he knew his "Work Classification" which is used to determine his salary. Besides by this information we thought that the perception of the opinion of the management will be influenced by the "Own Job Evaluation" as another source of information.

In diagram 1 we have represented the before specified hypotheses in a causal diagram. If the "Own Job Evaluation" does not correlate ($\gamma_{11} = 1$) perfectly with the "General Job Evaluation" the variable "Perception of the Evaluation by the Management" might explain the difference ($\beta_{12} \neq 0$). In this way we can test the effect of the perception variable on the answers of the respondents.

This model is tested for three qualities which have been evaluated: education, years of experience and leadership.

The respondents were asked to evaluate the job performance of a variety of hypothetical persons with different levels of formal education, years of experience and number of subordinates or leadership. For each topic the respondents had to evaluate the performance of the person with certain qualities compared with another as a standard. For the topic years of experience for example, was asked to evaluate the performance of persons with 1, 3, 10 etc. years of experience compared with a person with 5 years of experience. They were also asked to evaluate combinations of these qualities.

Subsequently the respondent gave an evaluation of his own qualities and he was asked how he perceived that his qualities were evaluated by the company. We thought that he could base his perception of the evaluation of his performance by the company on the "Classification" system of the company. This system was known to the respondent because his salary was determined by this classification system.

One of the stimuli, e.g. technical school, the interviewer presented to the respondent might have been respondent's own skill. In that case the respondent gave an evaluation of the respondent's own qualitities within a large set of evaluations of other qualities. But he also gave an evaluation of his own qualities we explicitly asked for.

We thougt that the first evaluation might be a more objective measure of his qualification in relation to other ones while the evaluations of respondent's own qualification in the explicit form might be biased by the respondent's perception of the opinions of relevant others.

In this case we expected biasing effects from the perception of the evaluation by the company as a relevant other because the respondent might expect some improvement of his position by exaggerating his qualifications compared with those given by the company.

Methodology.

A random sample of 505 workers from a Dutch steel factory constituted the respondents in this study. The variables were measured by magnitude estimation as developed by Stevens (a), introduced in the social science by Hamblin. The procedure used here is fully described in Saris et.al.. During the interview we trained the respondents in the use of magnitude estimation by asking them to evaluate the size of surfaces of rectangulars. The interviewer asked them to make an ordinal judgement and subsequently a magnitude estimation.

The introduction for the training was as follows:

> "The interviewer will give you seven cards on which figures have been drawn. They are numbered from A to F and one card has an S: compare the figure on card S with the other six figures.
>
> You can compare in words or signs by saying: this figure is very much bigger (+++), much bigger (++), bigger (+), equal (0), smaller (-), much smaller (--) or very much smaller (---).
>
> Then compare the figures using numbers. If we say that the standard (S) is 100, assign a number of 300 if a figure is 3 times as big. If the figure is ¼ as big as the standard assign the number 25".

After the respondent had read this instruction the interviewer put the standard card in front of the respondent. Then he shuffled the other cards and put the top one next to the standard. The respondent made first the verbal comparison and then a judgment in numbers. These judgments were written down by the respondent. The interviewer then took away the first stimulus to present the next one for comparison with the standard, and so on.

When the respondents understood this procedure the interviewer went on to the actual evaluations. For formal education it was done with the following words:

> "Now you have to mark in the same way how much education different persons have. You again make an evaluation in words or signs and in numbers. The standard (S) is a person who has finished technical school and two company organized courses".

All qualities were evaluated in this manner.

The respondent also evaluated his own qualifications in the same way and

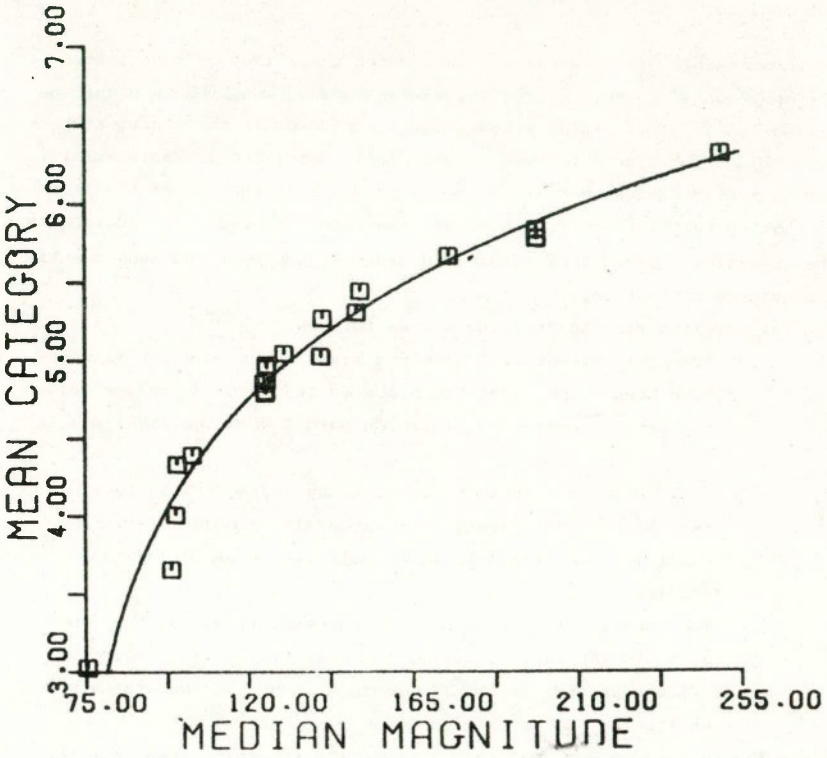

Figure 1. The relationship between the categorical evaluations and magnitude estimations for complex descriptions [1] based on aggregated data.

finally we asked him about his perception of his management's evaluation of his own qualifications.

As noted each evaluation is made once on a category scale and once on a metric scale. Because there are two measures it is possible to estimate the reliability if they represent the same variable (Jöreskog, Lord and Novick).

The question therefore is whether these two measures represent the same variable in this study.

In several other studies it has been shown that the two variables are not exactly the same but one measures similarity and the other magnitude (Eisler, a, b; Shinn): this is the explanation for their logarithmic relationship. Others observe that the category scale is a bad magnitude scale (Stevens, b; Wegener). However this may be, we found indeed the expected logarithmic relationship for several job qualities (fig. 1).

Quality	Goodness of fit R^2
education	.9973
experience	.9824
leadership	.9943
situation [1)]	.9924
complex descriptions [1)]	.9912

Table 1. The correlation between the log transformed magnitude scale and the category scale on aggregated data.

Observing a nearly perfect linear relationship (table 1) for aggregated data after logarithmic transformation of the magnitude scores we have treated the category scale scores and the log-transformed magnitude scores as measuring the same variable.

This procedure makes it possible to estimate the reliability of the different measures and to test the measurement model which we have specified in diagram 2.

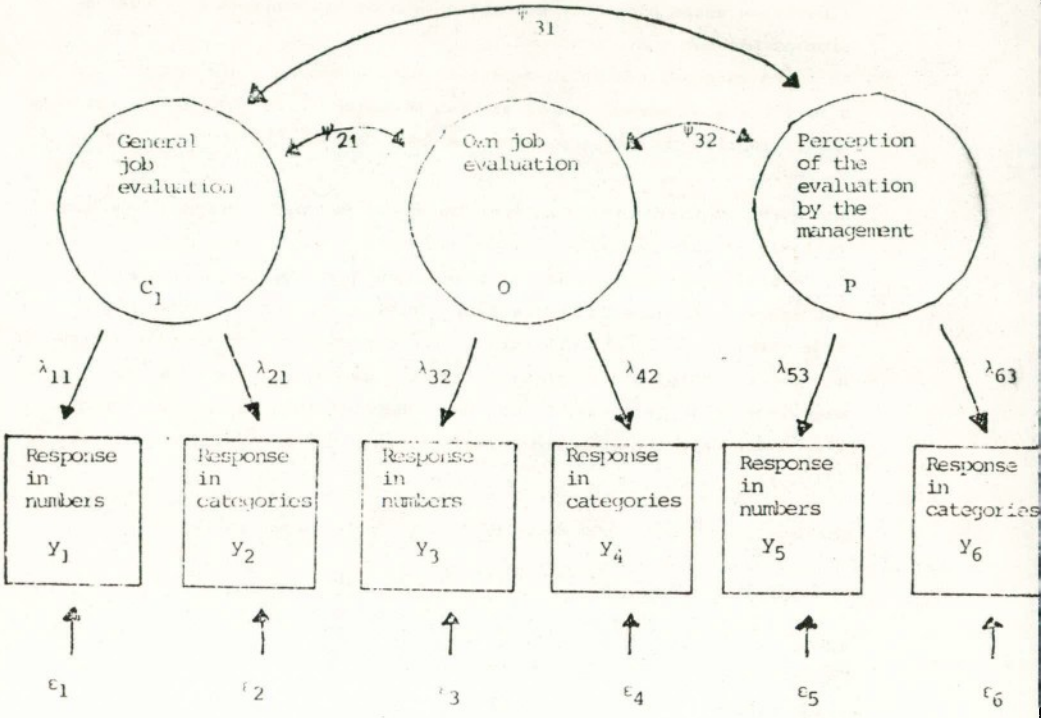

Diagram 2. A path diagram of the measurement model in this study.

In this model we do not mention the variable "Job Classification" as we had only one score of the measurement for this variable. In diagram 2 we have denoted the latent theoretical variables by C_1, O and P while y_1 to y_6 represent the observed scores for these variables. The measurement error in the observations is represented by ε_1 to ε_6 and ψ_{ij} represents the correlations between the theoretical variables which we want to explain according to the model of diagram 1. In the first step of this research we tested this measurement model, not making any restriction with respect to the correlations between the theoretical variables. This step is necessary because specification errors in this part of the model will also weaken the estimates in the remainder. In the second step the model of diagram 1 is tested and interpreted. As all models specified here are specific cases of the general LISREL system the data were analyzed by the program LISREL (Jöreskog and Sörbom). The correlation matrices for the three sets of job qualities have been given in appendix 1.

Results.

A. Analysis of the measurement model.

The results of the proposed analysis are summarized in table 2.

	χ^2_6	Probability
Experience	94.745	.000
Leadership	123.963	.000
Education	19.422	.003

Table 2. The goodness of fit of the model of diagram 2 for each data set

Table 2 indicates that the measurement model does not fit the data. We started to revise the model for the job quality experience. The first revision which seemed plausible was the introduction of covariances between the error terms of the categorical evaluations because it is easy for the respondent to remember these evaluations from one evaluation to the next. Therefore one might expect that the responses on the category scales are more correlated than might be expected from the correlations between the theoretical variables alone. These correlations are denoted by $\theta_{\varepsilon 42}$, $\theta_{\varepsilon 62}$ and $\theta_{\varepsilon 64}$.
The results of the test of this model was that $\chi^2_3 = 39.294$ and the Probability = .000.
The revision does give a significant reduction of the χ^2 value but this model still does not fit the data.
The next revision of the model was the introduction of a covariance between the error terms of the variables y_5 and y_3. This was done because in the interview the two questions were asked directly after each other which makes even for number responses memory effects possible if not likely. This correlation is denoted by $\theta_{\varepsilon 53}$.
The result of the test of this model was that $\chi^2_2 = 1.388$ and the Probability = .512. Thus this model fits the data very well for these job qualifications.
The same model was subsequently also tested for the other two job qualifications [2].
Table 3 shows the results of the test of this model on all three data sets.

	χ^2_2	Probability
Experience	1.338	.512
Leadership	1.548	.461
Education	1.363	.505

Table 3. The results of the test of the model of diagram 2 extended with the parameters $\theta_{\varepsilon 42}$, $\theta_{\varepsilon 62}$, $\theta_{\varepsilon 64}$, $\theta_{\varepsilon 53}$.

Clearly, the last model fits all three data sets.
Having found a fitting model, we can consequently say something about the random errors in this study.

1. This study clearly shows that it is hard to repeat similar questions using categorical scales: the memory effects are considerable.
2. The memory effects are less strong for magnitude estimate scales unless questions are asked directly in sequence.
3. Table 4 indicates that there is no clear difference in reliability of category and magnitude scales.

	experience	education	leadership
λ_{11}	1.000	.905	.977
λ_{21}	.814	.855	.817
λ_{32}	.668	.861	.854
λ_{42}	.859	.924	.861
λ_{53}	.730	.691	.874
λ_{63}	.953	.883	.779

Table 4. The square root of the reliability coefficients of the different observed variables.

B. The substantive model.

The correlations between the general job evaluation scores and own job evaluation scores are for Education: .581; for Experience: .765; and for Leadership: .616.

This indicates that the evaluation of one's own job is not identical to the general job evaluation: the correlations are quite different from 1 even after correction for measurement error as we have done by use of Lisrel.

It is hypothesized that in this case the perception of the evaluation of the management could be the explaining factor for the differences.

In our analysis the relationship between "Job Classification system" (C_2) and "Perception of Evaluation by Management" (P) is zero in this model. This indicates that the effect of C_2 on P is not distinquishable from the effect of "General Job Evaluation" (C_1) on P.

Accordingly, the evaluation seems to be independent of the factual information from the C_2 variable. Consequently the variable "Job Classification system" is ignored in the other analyses. [3)].

This leads to a model in which there are only effects of "General Job Evaluation" (C_1) and "Perception of Evaluation by Management" (P) on "Own Job Evaluation" (O).

Table 5 presents the results of the analysis using the measurement model as previously developed combined with the new model.

	Experience	Leadership	Education
χ^2_3	1.350	1.661	3.007
Probability	.717	.645	.390
γ_{11}	.788 *	.685 *	.593 *
β_{21}	.631 *	.829 *	.690 *
β_{12}	-.046	-.131	-.037

Table 5. The results for the test of the model of diagram 1 combined with the adjusted model of diagram 2.
* = significantly different from zero on the 5% level.

This model fits the data very well. Looking at the parameter values we see that the effects of "Perception of Evaluation by Management" (P) on "Own Job Evaluation" (O) are not significant. Therefore the model is further simplified assuming that the effects of the perception variables are zero. Table 6 summarizes the test of this model on the data sets of "Experience", "Leadership" and "Education".

	Experience	Leadership	Education
χ^2_4	1.539	2.064	3.043
Probability	.819	.723	.550
γ_{11}	.768 *	.620 *	.578 *
β_{21}	.610 *	.785 *	.671 *
β_{12} **	.000	.000	.000

Table 6. The results for the simplified model.
* = significant on the 5% level.
** = fixed at zero.

This latest model fits the data also very well. For all three job qualifications the χ^2_4 is only slightly higher and the probability is still very high.

The results of this analysis lead to the following substantive conclusions:

1. The general evaluation of aspects of one's own job has an effect on one's own job evaluation. But the relationship is not so high that they can be seen as measures of the same variable.
2. This study does not show an effect of "Perception of Evaluation by Management" on "Own Job Evaluation".
3. Perception of other's opinions is not directly influenced by obvious outside information such as a Job Classification system.
4. "Own Job Evaluation" has a strong effect on "Perception of Evaluation by Management". The sizes of these effects suggest that perceptions are mainly determined by internal processes.

Discussion.

This study has shown that at least for the "Evaluation of Job Performance" the effect of "Perception of Evaluation by Management" on the answers of the respondents is not significant. Consequently, it is not possible that in this case interviewer characteristics have an effect on the answers via these perception variables. As we can not imagine another mechanism which could produce such effects we think that these effects do not exist in this case. Although it is not possible from this study to generalize to all possible studies, we would like to express our doubt about such effects in general.
The published evidence could be a highly biased sample from all studies which have been done.
A systematic study of Heynen and Hagenaars support this point.
In this study the number of significant interviewer effects is not larger than the number which can be expected by chance.
Only a different publication policy could shed some light on this point. Publication of rejected hypotheses is necessary as long as the studies are correctly done.
Do we then believe that there are no systematic errors in survey data ?
In fact, in our analysis we have shown that the three "General Job Evaluations" and the three "Own Job Evaluations" are only correlated .581, .616 and .765 which is rather low if we consider that the stimuli were the same and that corrections for attenuation have been made.
Systematic errors therefore exist. The conclusion should be that "General Job Evaluation" is really different from "Own Job Evaluation" but what causes these systematic difference is another problem.
We still think that the process which produces "General Job Evaluation" is quite different from the mechanism which produces "Own Job Evaluation" especially because of personal involvement. But the perception of the opinion of management does not cause these differences. Further studies should be done to discover other biasing variables. As long as such information is not available we think that "General Job Evaluation" is a more valid procedure to obtain evaluations of performances on job qualifications.
An interesting finding of this study was that categorical judgments can not be repeated without relatively strong memory effects. This suggests the usefulness of a more wide spread application of techniques like magnitude estimation which are not so sensitive to this kind of problems.

Notes.

1) The respondents were also asked to evaluate their work circumstances and combinations of all qualifications combined in the same way. In this table they have been denoted by resp.: situations and complex descriptions.

2) We also tested a number of alternative models on all three data sets but the solution presented here was the only satisfactory solution for all three data sets.

3) Omitting this variable in the model would lead to an underidentified model if the disturbance terms were assumed to be correlated. If it is assumed that this correlation is zero the model is identified. That this assumption does not harm the fit is tested. For the model with this correlation as a free parameter the χ^2 test can still be used as test of goodness of fit although the model is not identified.
It turned out that in these cases the improvement of the fit is only very minimal and nonsignificant.
Therefore it is justified to fix this correlation at zero.

Appendix 1.

Zero order correlations for the data set Experience.

	y_5	y_6	y_1	y_2	y_3	y_4
y_5	1.000					
y_6	-.579	1.000				
y_1	.558	-.371	1.000			
y_2	-.360	.612	-.675	1.000		
y_3	.518	-.661	.345	-.464	1.000	
y_4	-.424	.753	-.255	.490	-.819	1.000

Zero order correlations for the data set Formal Education.

	y_5	y_6	y_1	y_2	y_3	y_4
y_5	1.000					
y_6	-.793	1.000				
y_1	.440	-.440	1.000			
y_2	-.504	.600	-.617	1.000		
y_3	.449	-.477	.224	-.314	1.000	
y_4	-.417	.573	-.257	.378	-.774	1.000

Zero order correlations for the data set Leadership.

	y_5	y_6	y_1	y_2	y_3	y_4
y_5	1.000					
y_6	-.740	1.000				
y_1	.735	-.574	1.000			
y_2	-.534	.745	-.674	1.000		
y_3	.509	-.516	.443	-.374	1.000	
y_4	-.433	.587	-.348	.421	-.798	1.000

References.

Dijkstra, W. and E. Elsinga, "Some determinants of interview bias", In: Methoden en Data Nieuwsbrief van de sociaal wetenschappelijke sectie van de V.V.S., 4, 1979, 91-118.

Eisler, H., The connection between magnitude and discrimination scales and direct and indirect scaling methods. In: Psychometrika, 1965, 30: 271-289.

Eisler, H., Empirical test of a model relating magnitude and category scales. In: Scandinavian Journal of Psychology, 1962, 3: 88-96.

Heinen, A.G. and J.A.P. Hagenaars, Research to Interviewer effects existing data. In: Methoden en Data Nieuwsbrief van de sociaal wetenschappelijke sectie van de V.V.S., 4, 1979, 123-130.

Henson, R., C.F. Cannell and S. Lawson, Effects of interviewer style on quality of reporting in a survey interview. In: Journal of Psychology, 1976, 93: 221-227.

Hetebrij, M., The Interview and Theory construction. Internal Report, Free University, Amsterdam, 1975.

Hyman, H.H., W.J. Cobb, J.J. Feldman, C.W. Hart and Ch.H. Stember, Interviewing in Social Research. Chicago, 1954

Johnson, R.F.Q., Pitfalls in research: The interview as an illustrative model. In: Psychological Reports, 1976, 38: 3-17.

Jöreskog, K.G. & D. Sörbom, LISPEL IV: A general computer program for estimation of linear structural equation systems by maximum likelihood methods. Chicago, 1978.

Lord, F.M. and M.R. Novick, Statistical Theories of Mental Test Scores. London, 1968.

O'Muircheartaigh, C.A., Response errors in an attitudinal sample survey. In: Quality and Quantity, 1976, 10: 97-115.

Phillips, D.L., Knowledge from what ? Theories and methods in social research. Chicago, 1971.

Phillips, D.L., Abandoning method. London, 1973.

Saris, W.E., C. Bruinsma, W. Schoots and C. Vermeulen, The use of magnitude estimation in large scale survey research. In: Mens en Maatschappij, 1977, 4: 369-395.

Shinn, A.M. Jr., Relations between Scales. In: Measurement in the Social Sciences. Theories and Strategies, H.M. Blalock, Jr. Chicago, 1974.

Stevens, S.S., Psychophysics: Introduction to its Perceptual Neural and Social Prospects. New York, 1975.

Sudman, S. and N.M. Bradburn, Response effects in surveys. Chicago, 1974.

Wegener, B., Das exponenten Problem beider psychophysischen Skalierung "sozialer" Variabelen. Report of Zentrum für Umfragen, Methoden und Analysen, Mannheim, 1978.

Weiss, C.H., Interaction in the research interview; the effects of rapport on response. Proceedings of the social statistics section of the American Statistical Association, 1970, 17-20.