# Social desirability bias, response order effect and selection effects in the new Dutch Safety Monitor

**Thomas Kraan, Jan van den Brakel, Bart Buelens and Harry Huys**

Statistics Netherlands.
e-mail tc.kraan@cbs.nl

***Abstract:***
*Recently, there has been a transition in the survey that measures perceived and actual safety in the Netherlands. This transition has caused significant discontinuities, especially for attitude questions related to the perception of the neighbourhood and the evaluation of the police performance. Different influences of the response order effect and social desirability bias and their interaction may explain these.*

*Before the transition, the Safety Monitor (SM) was conducted annually with a net sample size of about 20,000 respondents nationwide. Parallel to the SM local authorities collected data on the same topics at a regional level, which gave rise to inconsistent regional and national data about safety feelings and crime victimization. Because of this, the SM and the regional surveys were combined into one new survey, the Integrated Safety Monitor (ISM). Apart from other differences between the methodologies of the two surveys, the ISM employs Web Survey and Self Completion Paper Questionnaires in addition to Computer Assisted Personal Interviewing and Computer Assisted Telephone Interviewing as used by the SM. In autumn 2008, the ISM was performed for the first time with a net sample size of 62,803 respondents.*

*To assess the effect of the modified methodology, the SM has been executed in parallel with a limited net sample size of 6,113 respondents. Comparison of ISM with SM results reveals that the survey transition has caused discontinuities in some key safety indicators that can be traced back in part to the different sets of response modes employed in the two surveys and their related effects.The role of selection effects may be limited*

## 1. Introduction

Influences of survey modes on the survey outcomes have been studied for a long time (see e.g. Krosnick, 1999, for a review). With the introduction of mixed mode surveys, including internet modes, new mode effects and interactions with other factors like the questionnaire design arise. A well-known effect is the response order effect (see e.g. Krosnick and Alwin, 1987, for a review). The order in which response categories are offered to respondents affects which options are chosen, resulting in a preference for the first option for visual presentation (primacy effect) and a preference for the last option for oral presentation (recency effect). Cognitive burden plays a role in the theoretical explanation of this effect and the difference between visual and oral presentation. Another effect is that the presence of an interviewer elicits respondents to respond in a socially more desirable way (Krosnick, 1999, and references therein). Differences exist between response modes with respect to how the situation is dealt with when respondents cannot or do not want to make a choice. The 'don't know' option is generally not read aloud but can be selected by the interviewer in the case of interviewer administered questionnaires, but is often a valid option for the respondent for self-administered questionnaires. Mode effects as mentioned above occur frequently with attitude questions. In addition to these phenomena, in mixed mode surveys there can be selection effects, i.e. different subpopulations respond via different modes, each having their own distortion due to the mode effect. Thus, also selection effects can

Remarks: The views expressed in this paper are those of the authors and do not necessarily reflect the policies of Statistics Netherlands.

introduce a bias.

Recently, the Dutch victimization and safety survey was redesigned. The aforementioned effects may be responsible for discontinuities due to this transition since it was accompanied by the introduction of response modes not used before.

## 2. The Dutch Safety Monitor

In the Netherlands, the Safety Monitor (SM) measured actual and perceived safety. In this nationwide survey, respondents were asked to answer questions related to e.g. feelings of safety, opinions about police performance, and crime victimization. In the past few years (2006-2008) the SM was performed nationwide with a net sample size of about 20,000 respondents annually. This amounts to roughly 750 respondents in each of the 25 police districts. (The Netherlands is divided in 25 police districts). Local parties conducted their own independent surveys in order to produce reliable figures on a local level. Because of this independence, these local figures could not be compared mutually, nor could they be aggregated consistently on a national level. To overcome these issues, a new monitor has been introduced, the Integrated Safety Monitor. This monitor has a nationwide part but in order to allow the police districts and other local parties like municipalities to participate on a larger scale, the nationwide survey has been adapted thoroughly. In this new set-up local parties can increase sampling fractions among specific regional and local areas like police districts or neighbourhoods, in addition to the nationwide survey.

The questionnaire has been adapted, too. In the SM, all respondents were presented the questions on all topics. The new ISM questionnaire however has a modular design, consisting of obligatory and optional questionnaire blocks on specific themes, in a fixed order. In the nationwide sample, all respondents are presented all questions; in the local samples, local authorities are free in the choice of optional blocks. However, they can only choose complete blocks and not separate questions. New questions have been added in the ISM, but many questions are the same as in the SM. In some cases, there are slight changes in wording or answer categories. As an aside, with the transition from SM to ISM, also the data collection period shifts, from winter/spring to autumn.

In the weighting procedure of the ISM, national and local data are combined, and only one set of weights is obtained that is used to compute figures at both nationwide and regional levels. By using one single set of weights and a single questionnaire set, consistency between national and regional figures is guaranteed. In this manner, the inconsistencies occurring in the past (when regional parties conducted their own local survey with different questions independently) have been overcome. Thus, the ISM is an integrated measurement instrument that combines national and regional data on the same issues.

As mentioned above, apart from the geographical aspect, the questionnaire design, and the time frame for data collection, another difference between ISM and SM is in the use of response modes. The SM employed the Computer Assisted Personal Interviewing (CAPI) and Computer Assisted Telephone Interviewing (CATI) modes; the ISM employs in addition to the latter two, Web Survey (WS) and Self Completion Paper Questionnaire (SCPQ).
In the SM, people in the sample were approached via CATI if their phone number was known, otherwise CAPI was used. In the ISM, people in the sample are asked first to complete a questionnaire via WS. Alternatively, these people can respond via SCPQ. If neither of these modes are used, they are approached via CATI if their phone number is known, and via CAPI otherwise.

Since the two surveys, SM and ISM, employ different response modes and each mode has its specific effects affecting the response data, different results for the two surveys are to be expected. In order to assess the possible systematic effects of the survey transition between SM and ISM, both surveys were conducted in parallel. For this purpose an additional SM was executed, but on a smaller scale than the regular SMs in previous years.
In this paper, first possible significant differences between SM and ISM results are examined by comparing these for a limited set of variables, and testing the differences for their statistical significance. Second, the response data sets are explored to investigate by what mechanisms these differences may be caused.

## 3. Experiment

To analyse discontinuities in the main target parameters due to the changeover from the SM to the ISM, both surveys were conducted in parallel in the autumn of 2008. As both surveys are performed in the same period, neither seasonal effects nor trends can disturb the comparison. The samples for ISM and SM are drawn independently.

The sample for the ISM is drawn in two parts. First, there is a nationwide sample with a targeted number of 680 respondents per police district. This sample is drawn by means of stratified simple random sampling without replacement, where police regions are the strata. To achieve the targeted number of respondents per police region, the initial sample for the nationwide survey amounted to 24,966 elements and an additional 3,383 questionnaires have been sent out when there was the threat of the number of targeted respondents not being reached. Second, there was additional local sampling on request of local authorities for various geographical levels. For the local surveys, the sample amounted to 127,552 elements. As a result, the final sample can be considered as the realisation of stratified simple random sampling where the strata have a complex detailed regional classification.

The SM is based on stratified simple random sampling with police districts as strata, with proportional allocation. The targeted number of respondents for this survey amounts to 6,000. For this purpose, the sample size amounted to 9,167 elements.

In autumn 2008, the ISM and SM have been conducted as described above. The response data set for ISM thus obtained consists of national and regional data and has a net sample size of 62,803 respondents. Data collection for ISM was done by both Statistics Netherlands (SN) and external (non-SN) parties. The SM has a net sample size of 6,113 respondents. For this survey, data collection was performed by SN.

Parameter estimation for ISM and SM was then performed via the generalized regression estimator (Särndal, Swensson and Wretman, 1992) compensating, at least partially, for selective non-response. The weighting scheme for both surveys is based on a classification of different sociodemographic variables, see Buelens and Van den Brakel (2009) for the ISM and Integrale Veiligheidsmonitor (2008) for the SM.

Statistical inferences on the differences between the results of the two surveys is based on the design-based analysis procedure for experiments embedded in complex sampling designs, developed by Van den Brakel and Renssen (2005). In the present situation, their general approach reduces to a design-based *t*-statistic that accounts for the applied sampling design and weighting procedure of the ISM and SM.

## 4. Results

### 4.1 Discontinuities

The discontinuity as a consequence of the survey redesign has been determined for some key indicators, describing safety in general terms. It turns out that there is a significant discontinuity for most of these variables. Three indicators derived from attitude questions are considered. These indicators are based on a number (4 or 5) of underlying questions addressing various aspects of the quantity probed (e.g. for the quantity 'degradation of the environment', there is an underlying question on the occurrence of graffitti, as explained later in more detail). Each question gives an equal maximum number of score points, such that the range of the indicators is 0-10. The following indicators are considered in the sequel: perceived degradation of the neighbourhood (0: degradation doesn't occur, 10: degradation occurs frequently); perceived harassment of people in the neighbourhood (0: harassment doesn't occur, 10: harassment occurs frequently) and evaluation of police performance (0: negative evaluation, 10: maximally positive evaluation). The questions related to these three indicators are offered to all respondents. In addition, to those people in the sample who have had contact with the police on the occasion of an incident, a simple question on a 5-item scale is offered, probing (dis)satisfaction with the police performance after this contact. Thus, percentages of people (dis)satisfied with the police performance are computed. The percentage of these people having had contact with the police is another quantifier for safety. Table 1 presents for the six indicators mentioned, the results for ISM and SM with their sampling errors as well as the value of the *t*-statistic.

**Table 1: Aggregate scores for some attitude questions for SM and ISM**

| | SM | | ISM | | t-statistic |
|---|---|---|---|---|---|
| | score | sampling error | score | sampling error | |
| harassment in neighbourhood | 1.34 | 0.02 | 1.65 | 0.02 | 10 |
| police performance | 5.88 | 0.03 | 5.50 | 0.02 | 10 |
| degradation of neighbourhood | 2.97 | 0.03 | 3.64 | 0.02 | 18 |
| contact with police (%) | 28 | 0.6 | 31 | 0.4 | 2.8 |
| satisfied with police performance(%) | 55 | 1 | 62 | 0.8 | 6.7 |
| dissatisfied with police performance (%) | 29 | 1 | 21 | 0.7 | 7.7 |

The indicator for police performance given here for ISM is computed in a different way than in the official ISM figures, such as to allow comparison with SM. (The number of response items for the underlying questions differs for the two surveys: ISM: 5 items (e.g. ‘fully disagree’, ‘disagree’, ‘neutral’, ‘agree’, ‘fully agree’, adding resp. 0.0; 0.5; 1.0; 1.5; 2.0 score points to the indicator), SM: 3 items (e.g. ‘disagree’, ‘neutral’, ‘agree’, adding resp. 0, 1, 2 scorepoints to the indicator). The ISM results presented here are recalculated according to the SM 3-item division, by collapsing ‘fully disagree’/‘disagree’ (0 score points) and ‘agree’/ ‘fully agree’ (2 score points)).
Table 1 shows that for these six indicators the discontinuity as a result of the survey transition is significant at the 5% level. In the following subsections, it is attempted to explain these discontinuities.

### 4.2 Different answers via different modes

Since ISM and SM use different response modes, a mode dependency may at least partially explain the discontinuities between SM and ISM. To render evident this dependence, the response behaviour for the various modes is given for ‘perceived graffiti occurrence’ (Table 2) for the ISM. The score for this question contributes to the indicator ‘degradation of the neighbourhood’.

**Table 2: Response distribution (%) for perceived graffiti occurrence in the ISM**

| | CAPI | | CATI | | SCPQ | | WS | WS |
|---|---|---|---|---|---|---|---|---|
| | all | valid | all | valid | all | valid | all | valid |
| Graffiti | | | | | | | | |
| Occurs frequently | 12 | 13 | 10 | 10 | 12 | 13 | 11 | 12 |
| Occurs sometimes | 23 | 24 | 23 | 24 | 31 | 35 | 33 | 36 |
| (Almost) never occurs | 61 | 63 | 65 | 66 | 45 | 52 | 48 | 52 |
| Refuses | 0 | | 0 | | 1 | | 0 | |
| Don't know | 4 | | 1 | | 12 | | 7 | |
| Total | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

In this example, a shift is evident from the third response category ((almost) never occurs) to the middle category (occurs sometimes) for the SCPQ and WS modes as compared to the CAPI and CATI mode. A possible explanation is that for the self-administered questionnaires (SCPQ and WS), the respondents do not take time to consider and tend to opt for the middle category (a phenomenon called satisficing, Krosnick and Alwin, 1987). For the first category (occurs frequently) such a shift – as could be expected because of the primacy effect– does not occur: the percentages are roughly equal for all modes. It is also known that if response options are read aloud by an interviewer (as in CAPI and CATI), the respondent tends to choose for the last option (recency effect). Thus, the number of uncritical respondents under CAPI and CATI is enhanced, given the response order of this question. Both effects effectively result in a larger number of more critical responses under the SCPQ and WS modes: more people choose for the more critical middle category. Moreover, less people under the SCPQ and WS modes – as compared to CAPI and CATI– tend to choose for the last and uncritical option: the recency effect does not occur for the SCPQ and WS modes. It is also seen from Table 2 that under these self-administered modes more people respond 'don't know/refuses'. As indicated in section 1, the option 'don't know/refuses' is more readily available under these modes, making it easier to opt for it for respondents under these modes if they want to reduce their efforts to complete the questionnaire. Presser and Schuman (1989), Gilljam and Granberg (1993) and Van den Brakel et al., (2006) found similar results. Next to mode effects, selection effects can play a role here. We will return to this in subsection 4.4.

The previous example shows that mode effects seem to be present. They may provide an explanation for the observed differences for the six indicators. The effect mentioned above of the order in which the response categories are read aloud by the interviewer (recency effect) causes more positive, less critical, replies for perceived harassment and degradation of the neighbourhood as for the attitude questions, the least critical answer is read aloud last under the CAPI and CATI modes. Hence, this option is chosen more often, by the recency effect. Since the recency effect does not have an influence for the WS and SCPQ respondents in the ISM, the results for harassment are more negative here. On the other hand, for the question about police performance the negative response options are read aloud last. Hence for the ISM one would expect a more positive assessment as a consequence of the contribution of respondents via WS and SCPQ, for whom the recency effect does not play a role. We will see this demonstrated later. There is another mode effect, social desirability bias. The presence of an interviewer elicits socially desirable, i.e. positive and less critical, responses (social desirability bias, Krosnick, 1999). This would lead to a more positive evaluation for SM (counting only positively biased CAPI/CATI respondents) than for ISM (which has a large amount of WS and SCPQ respondents for whom there is a smaller social desirability bias). Table 1 reveals this more positive evaluation for SM. The recency effect seems not to prevail over the social desirability bias for this indicator. This is in accord with what is known from the literature: social desirability bias is a strong effect and the response order effect is a weak effect that may even be absent (Krosnick, 1999).

For the attitude questions about (dis)satisfaction with police performance on occasion of the last contact, there is a more positive evaluation for ISM than for SM (Table 1). This is in contrast to the scale score for police performance in general, where we see a more negative evaluation. However, two factors inhibit direct comparison of these seemingly contradictory observations. First, the presentation of the question differs. Second, the scale score is based on all respondents, whereas the (dis)satisfaction percentages are based on scores only from people who have had contact with the police. Despite this, we will try to shed some light on this paradox. The first factor, the difference in response order, seems relevant here: the differences between ISM and SM may be explained via the recency effect in the CATI and CAPI modes. For the questions on (dis)satisfaction with police performance on last contact, negative response categories are presented last. As a demonstration of the recency effect, these categories are chosen more often under CAPI and CATI and - for the ISM - less than by WS and SCPQ respondents, for which the recency effect does not play a role. In addition, selection effects may play a role. We will discuss these in subsection 4.4. The relatively more positive response under WS and SCPQ leads to a larger satisfaction and smaller dissatisfaction in ISM.

To illustrate the mode effects mentioned above, the scale scores for the six indicators related to attitude questions are given per response mode in Table 3. The figures in this table are obtained with Hájeks ratio estimator for a domain mean (Hájek, 1971), where each mode is considered as a domain. See for example Särndal et al. (1992), section 7.4, for an expression.

**Table 3: Attitude question per response mode (ISM)**

| | CAPI | CATI | SCPQ | WS |
|---|---|---|---|---|
| harassment in neighbourhood | 1.8 | 1.3 | 1.7 | 1.9 |
| police performance | 5.8 | 6.1 | 5.3 | 5.1 |
| degradation of neighbourhood | 3.4 | 3.0 | 4.1 | 4.1 |
| contact with police (%) | 36 | 29 | 26 | 33 |
| satisfied with police performance (%) | 55 | 60 | 64 | 62 |
| dissatisfied with police performance (%) | 28 | 24 | 18 | 18 |

For modes without an interviewer (i.e. WS and SCPQ), scores for perceived harassment and degradation are higher (or equal in the case of harassment for the small number of CAPI respondents) and the score for police performance is lower than for modes with interviewer assistance. In other words, for these questions under these modes respondents tend to be more critical, as a demonstration of the social desirability bias.

The people having had contact with the police are presented a question on a five-item scale, of which the item offered first expresses large satisfaction and the item offered last expresses large dissatisfaction. In this way, the percentages of people (dis)satisfied with the police are computed. As conjectured before, by the response order effect, one expects the number of dissatisfied people to be larger and the number of satisfied people smaller under CAPI and CATI. This can be seen in Table 3. One may attribute this to a selection effect, e.g. younger people, who have a less positive opinion about the police, respond especially via WS. However, elderly people use SCPQ instead, having the same social desirability bias, which compensates for this. We will discuss this in subsection 4.4 in more detail.

In summary, the explanations for the discontinuities given above – based on especially the social desirability bias and to a lesser extent the recency effect - seem to find support in the figures of Table 3.

### 4.3 Recalculated ISM results for the CAPI/CATI respondents only

To illustrate further the importance of mode, the ISM results have been recalculated using only data from CAPI and CATI respondents, as used in the SM. This can be done by a new weighting procedure including only CAPI and CATI respondents. Alternatively, one could apply a domain estimator for the CAPI and CATI responses within the total data set of ISM as applied in subsection 4.2. The results of these recalculations are given in Table 4. As can be seen, the results hardly differ for the two methods of recalculation.

**Table 4: Scores for attitude questions in ISM, recalculated for CAPI-CATI respondents (SM modes)**

| | | ISM | | SM |
|---|---|---|---|---|
| | (all respondents) | (CAPI/CATI reweighted) | (domain CAPI/CATI) | (all respondents) |
| harassment in neigbourhood | 1.65 | 1.46 | 1.45 | 1.34 |
| police performance 3-pt | 5.50 | 6.01 | 6.01 | 5.88 |
| degradation of neighbourhood | 3.64 | 3.12 | 3.08 | 2.97 |
| contact with police (%) | 31 | 31 | 31 | 28 |
| satisfied with police performance (%) | 62 | 60 | 59 | 55 |
| dissatisfied with police performance (%) | 21 | 26 | 25 | 29 |

This table shows that the differences of the ISM results with those of the SM are smaller if for the ISM only the CAPI and CATI responses are used. Therefore, the new response modes used in the ISM give rise to other overall results than in SM. The mode effects seem therefore to be responsible for a large part of the discontinuity between ISM and SM. However, for most indicators, the differences between recalculated ISM results (only SM modes) and the SM results remain significant.

The observation that the two methods of computing, i.e. a reweigthing of CAPI/CATI only and the domain estimator, yield the same result is will play a role in the investigation of selection effects in subsection 4.4.

**4.4 Selection effects**

In the previous subsections, it was attempted to explain the discontinuity between ISM and SM results via various mechanisms. In addition to those, selection effects (i.e. different subpopulations respond via different modes) may play a role. One of the new modes is WS, which is offered in first instance to all people in the sample; if they are unable or unwilling to use the internet, alternative modes are available. It is well known that adoption of computers and internet is dependent on age (next to e.g. on education and income). Consequently, different age groups may be represented differently in the four modes used in the ISM. Table 5 shows per age class the distribution over the various modes. The first four columns with ratios show the distribution over the age classes per mode, in the last two the modes with and without an interviewer are combined.

**Table 5: Distribution over the response modes of respondents per age class**

| Age class | CAPI | CATI | SCPQ | WS | CAPI/CATI | SCPQ/WS |
|---|---|---|---|---|---|---|
| 15-24 | 0.23 | 0.12 | 0.03 | 0.12 | 0.14 | 0.11 |
| 25-34 | 0.22 | 0.10 | 0.05 | 0.16 | 0.12 | 0.14 |
| 35-44 | 0.21 | 0.19 | 0.07 | 0.21 | 0.20 | 0.19 |
| 45-54 | 0.16 | 0.20 | 0.12 | 0.21 | 0.19 | 0.20 |
| 55-64 | 0.10 | 0.17 | 0.23 | 0.19 | 0.15 | 0.20 |
| ≥65 | 0.09 | 0.22 | 0.49 | 0.10 | 0.20 | 0.17 |
| total | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

Evidently, different age classes have varying preferences for modes. However, the distribution over modes with an interviewer versus self-administered modes is roughly constant over the age classes (last two columns of Table 5). It is exactly this aspect of the presence of an interviewer that was held responsible for the discontinuities. So, given this approximate constancy of the ratio of respondents with and without interviewer assistance over the age classes, even if a specific age group has different opinions on a subject, their responses suffer from mode effects in more or less the same way as those of other groups. The contribution of the selection effect to the discontinuities may therefore be limited.

A possible selection effect could be e.g. that more critical respondents react via the SCPQ and WS modes. Then, one would expect a higher number of less critical respondents via the CAPI and CATI modes. This would result in lower scores for perceived harassment and degradation under CAPI and CATI for the ISM than for SM and a higher score for police performance under CAPI and CATI for ISM than for SM. Tables 1 and 3 indicate that this is not the case (at the 5% significance level). Therefore, the suggested selection effect does not occur. As a second counterargument against selection effects, the recalculation of ISM results (Table 4), including only CAPI and CATI respondents in the weighting gives almost the same results as a computation on the CAPI/CATI domain. In other words, the weighting that usually compensates for overrepresentation of specific demographic groups, does not change the results as compared to a domain estimator, from which we conclude that there already is an even representation. Actually, as mentioned, the indifference in the way of calculation occurs already with inclusion weights only and no further weighting with respect to demographic background variables.

Table 4 suggests that there might nevertheless be selection effects. Of course, if the mode effect had been the only effect, the recalculated figures for ISM including only CAPI and CATI respondents would equal those for SM. However, some significant differences remain. For perceived harassment and degradation of the neighbourhood, the CAPI and CATI respondents are more critical, but for the scale score for police performance the CAPI and CATI respondents are less critical. The following might explain these at first contradictory observations. Among the CAPI/CATI respondents of ISM and SM there are relatively slightly more elderly (≥65) people, unwilling or unable to use the internet, and not applying for a written questionnaire. This may be a very particular subgroup. These may have more critical opinions about their neighbourhood as well as a more positive opinion about the police. It turns out that the many CATI respondents of the eldest age class under ISM indeed have an opinion about the police that is significantly more positive than the CATI respondents under SM. For younger people, the differences between ISM (recalculated figures) and SM are not significant for this question.

For the percentage scores on (dis)satisfaction with the police performance, the differences (compare the last two columns of Table 5) between the recalculated results for the ISM, including only CAPI/CATI respondents and the results for the SM may also partly be due to a selection effect (specific modes reach specific groups of respondents). As described before, the CAPI/CATI respondents contributing to the recalculated ISM results may be mostly elderly people, having a more positive opinion about the police. It turns out that for this question, for the eldest age category under CAPI in ISM the percentage of people satisfied with the police is higher than in SM.

## 5. Conclusion

We conclude that mode effects (response order effect and the social desirability bias) may explain part of the discontinuities due to the survey transition from SM to ISM. Selection effects may also play a role, for example the limited use of the Web response mode by elderly people. However, there are indications that selection effects, though present, have a limited net effect. In the current situation, with respondents being able to select their preferred response mode, the attributions of observed patterns to the effects mentioned can only be made in a tentative way.

**Acknowledgment**

We would like to thank our colleagues from the SRS division of Statistics Netherlands, especially Ger Linden, and Judit Arends-Tóth from the Methodology Department of Statistics Netherlands.

## References

- Brakel, J.A. van den, and R.H. Renssen (2005), *Survey Methodology*, 31, 23-40.
- Brakel, J. van den, R. Vis-Visschers and J. Schmeets (2006), *Field Methods*, 18 (3), 321-334.
- Buelens, B and J.A. van den Brakel (2009), *Weging Integrale Veiligheidsmonitor 2008*, CBS report, DMH-2009-04-15-BBUS.
- Integrale Veiligheidsmonitor 2008, Landelijke rapportage, CBS report.
- Gilljam, M., and D. Granberg (1993), *Public Opinion Quarterly* 57, 348–357.
- Hájek, J. (1971), Comment on a paper by D. Basu, in *Foundations of Statistical Inference*, Godambe, V.P. and D.A. Sprott, eds., 236, Toronto: Holt, Rinehart and Winston.
- Krosnick, J.A. (1999), Survey Research, *Annual Review of Psychology*. 50, 537-567.
- Krosnick, J.A. and D.F. Alwin (1987), *Public Opinion Quarterly* 51, 201-219.
- Presser, S., and H. Schuman (1989), The management of a middle position in attitude surveys, in *Survey research methods*, Singer, E. and S. Presser, eds., 108–123, Chicago: University of Chicago Press.
- Särndal, C.E., B. Swensson, and J. Wretman (1992), *Model Assisted Survey Sampling*, New York: Springer Verlag.