

In: W.E. Saris & I.N. Gallhofer (eds.) (1988).  
*Sociometric Research. Vol. 2: Data Analysis.*  
New York: St. Martin's Press.

### 3 Artifacts in Mail Surveys: the Influence of Dillman's Total Design Method on the Quality of the Responses

E. D. de Leeuw and J. J. Hox

#### 3.1 THEORETICAL BACKGROUND

Dillman (1978) has developed a set of procedures, called the Total Design Method (TDM), to improve the quality of mail survey data. Most research on the efficacy of the TDM has been concerned with the response rate. Using the TDM Dillman (1978, p. 27) reports an average response rate of 70 per cent for samples from the general population of the United States. There is some evidence for the cross-cultural usefulness of the TDM. For instance, in The Netherlands, Hox, de Leeuw and Duijx (1984a) report a response rate of 71 per cent for a sample of the general population. Their study shows that, to achieve a response rate as high as this, the TDM has to be followed completely.<sup>1</sup>

A high response rate does not necessarily imply a high validity. The validity of survey research can be threatened by sample composition bias, due to self-selection of respondents (Rosenthal and Rosnow, 1975), and by response bias from several sources (Rosenthal and Rosnow, 1969; Philips, 1971; Dijkstra and van der Zouwen, 1982).

In personal interviews, respondents have been shown to differ from the general population on several variables ('t Hart, 1974). Mail surveys, also, suffer from self-selection of respondents. According to Dillman (1978, p. 190) the response on the last, certified mailing tends to restore the balance. This has been confirmed in the study by Hox, de Leeuw and Duijx (1984b) for the variables age, education and occupation. Respondents probably also differ from non-respondents on personality characteristics. Rosenthal and Rosnow (1975) list a number of differences between volunteers (for psycho-

logical experiments) and non-volunteers. Among other things volunteers are generally better educated, they have a higher social economic status (SES) and they are more sociable.

Two types of response bias can be distinguished: bias threatening the external validity of survey research, and bias threatening the internal validity of survey research. Internal validity refers to the approximate validity with which we infer that certain relationships exist within data; external validity refers to the approximate validity with which we infer that relationships within the data can be generalised (see Cook and Campbell, 1979, p. 37). Classical research on artifacts, both in sociology (see Phillips, 1971) and in experimental psychology (see Rosenthal and Rosnow, 1969), has generally focused on threats to the external validity. The internal validity of (non-experimental) social research can be endangered by problems of measurement and causal interpretation of the data; these have been the subject of psychometric and sociometric research. The purpose of this study is to incorporate problems of internal validity, and specifically measurement problems, in the field of artifact research.

Two well-known threats to the external validity of survey research are the influence of social desirability and respondent role-anticipations. In their review of the literature, Sudman and Bradburn (1974) report a consistent effect of social desirability on responses. Bradburn *et al.* (1978) have introduced a question-specific operationalisation of social desirability: perceived question threat. The influence of respondent role-anticipations on the results of a survey interview has been investigated by Dijkstra (1983), who reports some effects of experimentally induced role-anticipations on the responses. The internal validity of survey research can be threatened by partial non-response and deviant individual response patterns. Partial non-response is discussed by Nederhof (1981, p. 56); Bradburn *et al.* (1978) have investigated partial non-response in connection with question threat and response bias. An early discussion of individual response patterns can be found in Turner and Fiske (1968); more recent developments are discussed by Tatsuoka and Tatsuoka (1983).

Social desirability, question threat, role-anticipations, partial non-response and individual response patterns may all exercise an influence on the quality of the data in mail surveys. In mail surveys, using the Total Design Method of Dillman (1978), this influence may be enhanced by the growing pressure on potential respondents to react. This pressure is expected to be especially high for the last reminder, which is sent by certified mail, and which has a cover letter that is rather harshly worded.

### 3.2 RESEARCH DESIGN AND PROCEDURE

The subject of this study is to explore the effect of Dillman's Total Design Method on sample composition bias, and on response bias. Special consideration is given to the effect of the growing pressure, exercised by the reminders, on sociological and psychological attributes of the sample (Rosenthal and Rosnow, 1975) and on question threat, role-anticipations, partial non-response, and individual response patterns. The content of the questionnaire has been manipulated by using two different loneliness scales, developed by de Jong-Gierveld. The topic of one of these scales is rather common feelings of loneliness; the topic of the other scale is feelings of extreme social isolation. Both scales conform to the Rasch model (de Jong-Gierveld and Kamphuis, 1983).<sup>2</sup>

The attributes of the sample measured in this study are age, sex, education, research experience and town size. Sociability has been operationalised as a sub-dimension of extraversion; the other aspect of extraversion is impulsiveness (Fey, 1979). Sociability and impulsiveness were measured using a short version of the extraversion scale of the Dutch Five Personality Traits Test (Elshout and Akkerman, 1973). Social desirability was measured with a Dutch version of Crowne and Marlowe's Social Desirability scale (van Rooijen, 1981). Based on the study by Bradburn *et al.* (1978), a questionnaire threat scale has been developed; this scale conforms to the Mokken model (de Leeuw and Hox, 1984).<sup>3</sup> Three types of respondent role-anticipation – faithfulness, apprehensiveness and suspiciousness – were measured with a set of items by Dijkstra and Elsinga (1981). The deviance of the individual response patterns has been operationalised as a low scalability of the individual responses on the two loneliness scales. Since both loneliness scales conform to the Rasch model, the probability of individual response patterns can be calculated. In this study van de Flier's Q was used as an index of scalability (van de Flier, 1980).

Respondents have been grouped as to whether their questionnaire was a response to the first, the second, or the third mailing. The three mailing groups have been mutually compared on both sociological and psychological attributes, and on questionnaire threat, respondent role-anticipations, number of missing values and scalability of the loneliness scale. For each questionnaire the sample composition bias has been investigated by comparing the sample obtained with the general population, using figures from The Netherlands Central Bureau of Statistics (SCR, 1980; CBS, 1983). For the response effect

variables the population values are not available; only differences between the three mailing groups can be investigated. For the purpose of comparison the results of other studies using these variables are presented in the discussion section. Finally, the influence of questionnaire threat, respondent role-anticipations, and number of missing values on loneliness scores was investigated.

All effects reported in this study are significant on at least the 0.05 level.

Two types of questionnaires, which only differ in the first seven items, were used. One – the extreme version – contained the extreme loneliness scale; the other – the neutral version – contained the general loneliness scale. Using the telephone directory as the sampling frame, two random samples of the general population were drawn. The total number of addresses was 1000; 500 for each of the two samples. The complete Dillman method was used, including a last reminder by certified mail.

### 3.3 RESPONSE RATE AND INSTRUMENTS

The extreme version resulted in 264 returned questionnaires, 56 'undeliverable' questionnaires and 56 overt refusals. The response rate is 59 per cent. The neutral version has a response rate of 68 per cent: 301 returned questionnaires, 55 'undeliverables' and 61 overt refusals. The difference in response is statistically significant; there is no difference between the versions in overt refusals.

With the exception of the questionnaire threat scale, which has been developed specifically for this study, all scales can be found in the literature. The actual items and an extensive analysis of all instruments used in this study can be found in de Leeuw and Hox (1984). The most important results of this analysis are summarised in Table 3.1, and below.

The reliability of all scales is sufficiently high for the purposes of this study.

The items of the general loneliness scale and the extreme loneliness scale have been dichotomised using the same cutting-points as de Jong-Gierveld and Kamphuis (1983). After dichotomisation the scalability of the two loneliness scales is good: Loevinger's H is larger than 0.40, and inspection of the binomial plots for the Rasch model (Molenaar, 1983) reveals no anomalies.

The items of the Crowne and Marlowe Social Desirability scale and of the Questionnaire Threat scale are dichotomous. The scalability of

Table 3.1 Some psychometric properties of the instruments

| Scale       | n-items | reliab. ( $\alpha$ ) | Mokken (H) | Rasch |
|-------------|---------|----------------------|------------|-------|
| Lonely-G    | 7       | 0.77                 | 0.42       | yes   |
| Lonely-E    | 7       | 0.79                 | 0.41       | yes   |
| Sociable    | 5       | 0.75                 | –          | –     |
| Soc. Des.   | 20      | 0.60                 | 0.10       | x     |
| Quest. Thr. | 5       | 0.79                 | 0.59       | no    |
| Faithful    | 8       | 0.79                 | –          | –     |
| Anxious     | 5       | 0.81                 | –          | –     |
| Suspicious  | 6       | 0.76                 | –          | –     |

the Social Desirability scale is poor: the mean inter-item correlation is 0.07, and Loevinger's H is 0.10. No attempt has been made to test the Rasch scalability of the Social Desirability scale. The scalability of the Questionnaire Threat scale is good: the mean inter-item correlation is 0.43, and Loevinger's H is high. Rasch scalability of the Questionnaire Threat items is rejected by the Andersen ICCSL-test; inspection of the binomial plots (Molenaar, 1983) shows that the item-characteristic curves of the items are not parallel.

The items measuring respondent role-anticipations have been factor analysed. The factor structure reported by Dijkstra and Elsinga (1981) is replicated.

The Crowne and Marlowe Social Desirability scale has been removed from further analyses because of its low scalability. All other scales have been retained. The good scalability of the loneliness scales permits determination of the scalability of the individual response patterns of respondents for these scales. The scalability of the individual response patterns of the respondents for the Questionnaire Threat scale has not been analysed; this scale has not been used before, and further analyses await replication of the scalability of the items for this scale. To avoid complications arising from the use of two different loneliness scales, and the creation of variables which are experimentally dependent, the percentage of missing data is computed as the percentage of missing responses excluding the items of the loneliness scales.

### 3.4 RESULTS I: SAMPLE COMPOSITION BIAS

No differences in sample composition bias have been found between the two versions of the questionnaire.

Compared with the general population the sample is older (the

mean age is 45 years as compared to 41 years for the general population); it contains too many male respondents (56 per cent as compared to 50 per cent in the general population),<sup>4</sup> and it contains too many respondents from rural areas, and too few from small towns (the number of respondents from medium-sized and large towns does not differ significantly from the expectation). For these variables, there is no difference between the three mailing groups; as a consequence, follow-up mailings do not restore the balance in these cases.

Furthermore, the educational level of the sample differs from the educational level of the population; there are too few respondents with either a high, or a low, educational level. The educational level in the three mailing groups, which is shown in Table 3.2, differs significantly.

Table 3.2 Education of respondents by mailing group and in general population

| Education | Mail. 1<br>(N=390) | Mail. 2<br>(N=106) | Mail. 3<br>(N=69) | Tot.<br>(N=565) | Pop. |
|-----------|--------------------|--------------------|-------------------|-----------------|------|
| 1 Low     | 17.5               | 30.7               | 26.2              | 21.0            | 31.1 |
| 2 Middle  | 54.6               | 54.5               | 49.2              | 54.0            | 39.2 |
| 3 High    | 27.9               | 14.9               | 24.6              | 25.0            | 29.7 |

(All results significant at the 0.05 level.)

As Table 3.2 shows, the follow-up mailings bring the sample values closer to the population values, but the last (certified) mailing does not seem to be more efficient at this than the second.

For the variables 'Sociable' and 'Research Experience' no population values are available. The three mailing groups show a significant ( $p=0.05$ ), and systematic, difference for these variables; the results are given in Table 3.3.

Table 3.3 Volunteer characteristics of respondents by mailing

| Attribute | Mail. 1 | Mail. 2 | Mail. 3 | Mean | St. dev. |
|-----------|---------|---------|---------|------|----------|
| Sociable  | 14.2    | 14.9    | 15.5    | 14.5 | 4.12     |
| Res. exp. | 1.3     | 1.0     | 0.5     | 1.1  | 1.93     |

The last mailing groups are more sociable and have less research experience. According to Rosenthal and Rosnow (1975) volunteers are more sociable and have more research experience than non-volunteers. Our first mailing group is the least sociable of the three mailing groups, which conflicts with Rosenthal and Rosnow's findings. The number of research experiences of our first mailing group is almost equal to the number of research experiences Nederhof (1981, p. 100) quotes for a sample of the general population interviewed in 1979; the number of research experiences of the second and third mailing groups is significantly lower.

### 3.5 RESULTS II: RESPONSE BIAS

Again no differences in response bias have been found between the two versions of the questionnaire.

The attitudes of the respondents in the total sample towards the study differed for the three mailing groups:

Table 3.4 Mean respondent attitude by mailing

| Attitude    | Mail. 1 | Mail. 2 | Mail. 3 | Mean | St. dev. |
|-------------|---------|---------|---------|------|----------|
| Faithful    | 47.8    | 46.5    | 44.4    | 47.1 | 6.04     |
| Apprehens.  | 12.6    | 15.0    | 14.4    | 13.3 | 6.50     |
| Suspicious  | 19.8    | 21.6    | 21.8    | 20.4 | 7.38     |
| Quest. Thr. | 1.7     | 2.0     | 2.3     | 1.8  | 1.73     |

(All results significant at the 0.05 level.)

Respondents in the first mailing group have more positive attitudes towards the study than respondents in the second, and third, mailing group. However, contrary to our expectations, it is not true that the difference between the third mailing (which is sent by certified mail), and the second, is larger than the difference between the second mailing and the first.

Also, contrary to our expectations, no differences were found between the three mailing groups concerning the percentage of missing data, and the scalability of individual response patterns for the loneliness scales.

To explore the influence of response bias on the results of mail survey research, the correlations of the variables, which measure

various aspects of response bias with the two loneliness scales, are given in Table 3.5.

Table 3.5 Correlations between response bias variables and loneliness scales

| Scale/Var. | Faithf. | Appr. | Susp. | Q. Thr. | No. Mis. | Scalab. |
|------------|---------|-------|-------|---------|----------|---------|
| Gen. Lon.  | -0.08   | 0.11  | 0.08  | 0.07    | -0.01    | -0.02   |
| Extr. Lon. | -0.02   | 0.19* | 0.29* | 0.23*   | 0.21*    | -0.27*  |

\* significant at the 0.05 level.

The correlation pattern shows a difference between the two versions of the loneliness scale. For general loneliness the correlations between loneliness and the response bias variables are negligible. For extreme loneliness the correlations between loneliness and most response bias scales is significant, although not extremely high.

### 3.6 CONCLUSIONS AND DISCUSSION

Compared to the general population the sample obtained in this study is older, contains too many male respondents, too many respondents from rural areas, too few from small towns and too few respondents with either a high or a low educational level. With respect to educational level, the second and third mailing bring the sample values somewhat closer to the population values. The sample composition bias is similar to the sample composition bias found using personal interview methods ('t Hart, 1974; Jones and Lang, 1980). There is one marked difference; in our study men are over-represented, while studies using personal interview methods generally show an overrepresentation of women. This overrepresentation of men in a mail survey has also occurred in earlier studies using the Total Design Method (Hox, de Leeuw and Duijx, 1984b). The sample composition bias is influenced by the choice of the telephone directory as the sampling frame as well as by self-selection of respondents. However, since telephone penetration in The Netherlands is well over 90 per cent (Bronner, 1980), while response rates, even for face-to-face interviews, is usually at most 70 per cent, the latter seems to be the most important mechanism to explain sample composition bias. Especially the overrepresentation of male respondents can

easily be explained by self-selection, in this case due to the tendency of women to refuse co-operation ('t Hart, 1974; van Tulder, 1977, 1978; Hippler and Seidel, 1985); in interview surveys this is (over)-compensated by the fact that women are, in general, more often at home than men ('t Hart, 1974; Inter/View groep, 1978).

Sample composition bias which is caused by imperfections in the sampling frame can be compensated by weighting procedures. This is not true, however, for sample composition bias which is caused by a refusal to co-operate. In this case simple weighting will not remove bias in variables which are correlated with refusing co-operation. If a study uses mixed-mode data-collection within the same sampling frame, investigators may try to disentangle the different effects and estimate the impact of sample composition bias on the results.

Respondents in the first mailing group tend to be 'better' research subjects than respondents in the second and the third mailing group. While the later mailing groups did not produce more missing data, or deviant individual response patterns, than the first mailing group, respondents in the first mailing group did have more positive attitudes to the study; they also have more research experience. Since these variables are shown to be correlated with extreme loneliness, the possibility of contamination of the variables of substantive interest is present.

One objection can be raised to this argument about the contamination of the extreme loneliness scale with response bias. Respondents who feel extremely lonely may, actually, be more apprehensive, and suspicious, research subjects. This objection does not mean that there is no possibility for contamination, but rather that such a contamination may be limited.

The difference in respondents' attitudes between the three mailing groups may reflect the reaction of respondents to the increasing pressure exerted by the Total Design Method. It may also reflect negative attitudes toward survey research which already exist in the population, especially in those groups which are difficult to reach by mail survey methods. The design of the study permits both interpretations. However, if the difference between the three mailing groups reflects existing differences in the population, then within each mailing group there should be a correlation between the willingness to respond (as measured by the order in which the completed questionnaires were received), and the attitudes of the respondents. No such correlations were found. Furthermore, when we look at the data collected by Dijkstra and Elsinga (1981) after an extensive interview

on loneliness, we see that after this interview the mean scores on faithfulness, apprehensiveness and suspiciousness were respectively 51.1, 10.2 and 13.0, while we find overall means of 47.1, 13.3 and 20.4 after a mail survey (de Leeuw Hox, 1984). These two findings, taken together, make the interpretation that the difference between the mailing groups is due to the different treatment of the mailing groups by the Total Design Method more likely.

Two types of response bias have been distinguished: response bias which affects the external validity of mail survey research (perceived questionnaire threat and respondent role-anticipations) and response bias which affects the internal validity of mail survey research (deviant individual response pattern and percentage of missing data). A comparison of the three mailing groups shows a difference with respect to the first type of response bias and no difference with respect to the second type of response bias. The utility of the distinction between these two types of response bias can also be assessed by an examination of the correlations between the response bias variables. Table 3.6 gives the result of a factor analysis (principal components followed by varimax-rotation) of these correlations.

Table 3.6 Component analysis of response bias variables

|            | I     | II    |
|------------|-------|-------|
| Q. Threat  | 0.53  | -0.06 |
| Faithf.    | -0.65 | 0.13  |
| Suspicious | 0.67  | 0.15  |
| Anxious    | 0.69  | 0.21  |
| Pct. Mis.  | 0.02  | 0.73  |
| Scalab.    | -0.03 | -0.66 |

The factor pattern in Table 3.6 clearly supports the distinction between the two types of response bias. It is recommended that studies on response bias should include variables which tap both types of response bias. Both the percentage of missing data and the scalability of the individual response patterns may be utilised. A different approach to the analysis of individual response patterns is taken by Bruinsma and Saris (1984); how well this approach fits in with the variables utilised in this study remains to be investigated.

McGuire (1969) has pointed out that in the life of an artefact three stages can be distinguished: ignorance, coping, and exploitation. The

investigation of response bias which threatens the internal validity of social research seems to be in the second stage (coping). The existence and possible importance of the artefact are being assessed; what remains to be done in this stage is to devise procedures which will reduce its contaminating influence. From a theoretical point of view the third stage (exploitation) is the most interesting. In this stage the artefact begins to receive research attention, not as a contaminating factor that has to be coped with, but as an interesting variable in its own right (McGuire, 1969, pp. 15-16).

From a practical point of view there is no objection to the use of follow-up mailings by ordinary, or certified, mail in order to increase the response rate. However, the use of follow-up mailings also increases the risk of contamination of the results by response bias. Researchers should recognise this artefact and attempt to cope with it.

## NOTES

1. Important features of the TDM are: personalised cover letters, a simple and attractive questionnaire and the use of follow-up mailings. Three weeks after the first mailing all non-respondents receive a new questionnaire with cover letter; seven weeks after the first mailing this procedure is repeated, but this time the questionnaire is sent by certified mail (Dillman, 1978).
2. The Rasch model is a probabilistic generalisation of Guttman's scalogram which imposes some very restrictive assumptions on the data (Rasch, 1960).
3. The Mokken scale is a non-parametric probabilistic generalisation of Guttman's scalogram (Mokken, 1970).
4. Omitting age < 18 years and age > 70 years.

## REFERENCES

- BRADBURN, N. M., SUDMAN, S., BLAIR, E. and STOCKING, C., 'Question Threat and Response Bias', *Public Opinion Quarterly*, **42** (1978) 221-34.
- BRONNER, F., 'Telefonisch Onderzoek', *Methoden en Data Nieuwsbrief*, **5** (1980) 145-56.
- BRUINSMA, C. and SARIS, W. E., 'Effects of Response Functions in Survey Interviews on Evaluation of Job Qualifications', *Kwantitatieve Methoden*, **5** (1984) 87-109.
- CBS: Central Bureau of Statistics, *Statistisch Zakboek 1983* (Den Haag: Staatsuitgeverij, 1983).
- COOK, T. D. and CAMPBELL, D. T., *Quasi Experimentation: design and analysis issues for field studies* (Chicago: Rand McNally, 1979).

- DIJKSTRA, W., *Beïnvloeding van Antwoorden in Survey-interviews* (Delft: Bouwplatenwerkplaats, 1983).
- and ELSINGA, E., 'Respondent-rolopvattingen in Survey-interviews', *Mens en Maatschappij*, 3 (1981) 311–17.
- and ZOUWEN, J. van der (eds), *Response Behaviour in the Survey Interview* (London: Academic Press, 1982).
- DILLMAN, D. A., *Mail and Telephone Surveys* (New York: Wiley, 1978).
- ELSHOUT, J. J. and AKKERMAN, A. E., 'Een Nederlandse test voor vijf persoonlijkheidsfactoren, de 5 PFT', in DRENTH, P. J. D. and WOLFF, Ch. J. de (eds) *Arbeids – en organisatiepsychologie* (Deventer: Kluwer, 1973).
- FEY, J. A., *Temparement, een onderzoek naar de betekenis van extraversie, impulsiviteit en spanningsbehoefte* (Lisse: Swets, 1979).
- FLIER, H. van de, *Vergelijkbaarheid van individuele testprestaties* (Lisse: Swets, 1980).
- HART, H., 't *Selectie en zelf-selectie van informanten in enquetes*, Dissertation (University of Amsterdam, 1974).
- HIPPLER, H. J. and SEIDEL, K., *Schriftliche Befragung bei Allgemeine bevölkerungs-stichproben – Untersuchungen zur Dillman'schen 'Total Design Method'*, ZUMA-Nachrichten (Mannheim: ZUMA, 1985).
- HOX, J. J., LEEUW, E. D. de and DUIJX, A. W. M., 'The Postman Rings Thrice', *Mens en Maatschappij* (1984a). (An English version is available as: HOX, J. J., LEEUW, E. D. de and DUIJX, A. W. M., 'The Postman Rings Thrice: the effect of personalizing and using certified mail on the response in mail surveys' (Amsterdam: University of Amsterdam, Department of Education, 1984b).
- Inter/View groep, *Rondje Nonresponse* (Amsterdam: Inter/View groep, 1978).
- JONES, W. H. and LANG, J. R., 'Sample Composition Bias and Response Bias in a Mail Survey: a comparison of inducement methods', *Journal of Marketing Research*, 17 (1980) 69–76.
- JONG-GIERVELD, J. J. de and KAMPHUIS, F. H., 'The Development of a Rasch-type Loneliness Scale', in *Multivariate Analysis in the Social and Behavioural Sciences* (Amsterdam: SISWO, 1983) 65–91.
- LEEUW, E. D. de and HOX, J. J., *Artifacten bij postenquetes: de gebruikte instrumenten* (University of Amsterdam: Subfaculteit Opvoedkunde and Subfaculteit Sociologie, 1984).
- McGUIRE, W. J., 'Suspiciousness of Experimenter's Intent', in ROSENTHAL, R. and ROSNOW, R. L. (eds), *Artifact in Behavioral Research* (New York: Academic Press, 1969).
- MOKKEN, R. J., *A Theory and Procedure of Scale Analysis* (Den Haag: Mouton, 1970).
- MOLENAAR, I. W., 'Some Improved Diagnostics for Failure of the Rasch Model', *Psychometrika*, 48 (1983) 49–72.
- NEDERHOF, A. J., *Some Sources of Artifact in Social Science Research: nonresponse, volunteering, and research experience* (Dissertation, University of Leiden, Leiden, 1981).
- PHILLIPS, D. L., *Knowledge from What?* (Chicago: Rand MacNally, 1971).

- RASCH, G., *Probabilistic Models for Some Intelligence and Attainment Tests* (Copenhagen: Danish Institute for Educational Research, 1960).
- ROOIJEN, L. van, 'Problem Drinking among Students in a Dutch University', *Gezondheid en Samenleving*, 2 (1981) 65–77.
- ROSENTHAL, R. and ROSNOW, R. L., *Artifact in Behavioral Research* (New York: Academic Press, 1969).
- and — *The Volunteer Subject* (New York: Wiley, 1975).
- SCR. *Social and Cultural Report 1980* (Den Haag: Staatsuitgeverij, 1981).
- SUDMAN, S. and BRADBURN, N. M., *Response Effects in Surveys: review and synthesis* (Chicago: Aldine, 1974).
- TATSUOKA, K. K. and TATSUOKA, M. M., 'Spotting Erroneous Rules of Operation by the Individual Consistency Index', *Journal of Educational Measurement*, 20, 3 (1983) 221–30.
- TULDER, J. J. M. van, 'Op de grens van de non-response I', in *Jaarboek van de Nederlandse Vereniging van Marktonderzoekers* (Amsterdam, 1977).
- 'Op de grens van de non-response II', in *Jaarboek van de Nederlandse Vereniging van Marktonderzoekers* (Amsterdam, 1978).
- TURNER, C. B. and FISKE, D. W., 'Item Quality and Appropriateness of Response Processes', *Educational and Psychological Measurement*, 28 (1968) 297–315.