

chapter 6

Memory effects in MTMM studies

A. VAN MEURS AND W.E. SARIS

The MTMM approach to evaluate methods of data collection in attitude and opinion research is based on the repetition of questions on the same item with different methods. Such an approach allows estimates of validity, invalidity and reliability coefficients, which are very important in correcting for measurement error in substantive research. However, repeating questions about the same item can also have its drawbacks. Many respondents will try to be consistent in their answers across methods. They will of course succeed if they can remember their previous answers. Such memory effects can bias the estimates of the quality indicators for measurement instruments in an MTMM model. In this paper we will discuss the following topics:

- the possible effects of memory
- to what degree respondents are able to remember their previous answers
- the factors which affect memory
- the consequences for the design of MTMM studies.

The data discussed in this paper were collected in April 1989. The respondents (N=1537) of the NIPO-telepanel (de Pijper and Saris, 1986; Verwey, et al., 1989) answered two questionnaires with an interval of two weeks between their presentation. Three questions in the first wave concerned the attractiveness of the three main Dutch political parties measured using a 10-point category scale. The same response scale was used for the attitude towards six public services and three life satisfaction questions. After a number of other questions half of the sample (N=795) was asked if they could remember their initial answers to items 1 to 3 and to give their opinion again on items 4 - 6 using the line production

The authors would like to thank H. Seip for his contribution, and H. Hartman and K. de Rooy for their work on the survey design.

MEMORY EFFECTS IN MTMM STUDIES

method. The other half of the sample (N=751) was tested for their memory of the questions 4-6 and they were asked to express their opinion about the items 1-3 using the line production method. Two weeks later all of the respondents cooperating (1381) completed a second interview. Again two groups answered a memory question on three items (group 1 on items 7-9, group 2 on items 10-12). Furthermore, all respondents gave categorical judgements on all items and line productions for all questions not evaluated in that way. The complete design is summarized in figure 1. The chosen design allows the correlations and MTMM estimates to be obtained for all variables with a short interval between the observations (within the same interview) and with a two week interval between the observations.

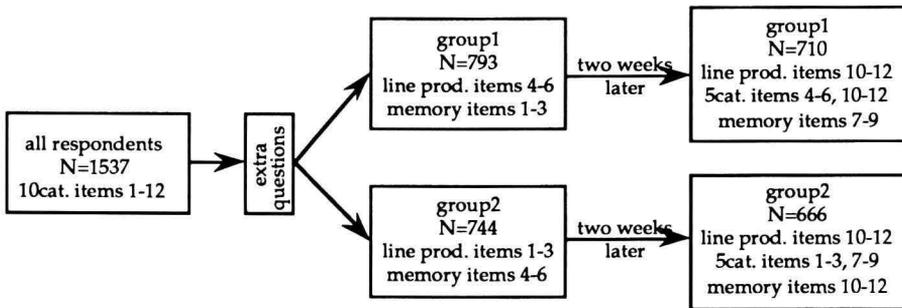


figure 1: Design of the study

THE POSSIBLE EFFECTS OF MEMORY

Repeating the same questions in a single survey can have its drawbacks. Some respondents are annoyed because repeating the same questions is considered as a waste of time. Others feel that they are not trusted, and that their answers are secretly checked by the researchers. These problems can be minimized by introducing the repeated questions with some sort of explanation. Our experience is that respondents then answer the questions seriously.

A more troublesome problem arises when respondents can remember their previous answers. If respondents are recalling their previous answers in the formulation of their reply to the repeated questions, then so called memory effects which are assumed to be zero in the MTMM model, might arise. If these memory effects are ignored, the estimates of the quality of the measurement instru-

CHAPTER 6

ments will be biased. This will generally lead to overestimation of the quality of the instruments, especially their validity.

Another possibility is that repetition of a question will give the respondents the impression that the questions are of some importance to the researcher and, consequently, they will think more carefully about their answers the second time. The answer could then change between the first and second time, thereby leading to lower correlations and lower estimates of the quality of the measurement instruments. This problem has been discussed in a set of papers by Jagodzinski et al. (1987) and Saris and van den Putte (1988) as the so-called "socratic effects" of asking questions.

From this brief discussion it is not clear what the expected effect will be in practice for the size of the correlations and the MTMM parameters. In the experiment described we could estimate the different effects on the correlation and the parameter estimates of the MTMM model after an interval of approximately 9 minutes (on average) within an interview and after an interval of two weeks in two different waves of the same panel. The results for the correlations are summarized in table 1.

table 1: Correlations between two methods with different time intervals

methods: time interval:	10 cat-line		5 cat-line	
	±9 min.	2 weeks	±9 min.	2 weeks
1 PvdA (labour)	.851	-	-	.824
2 CDA (chr. dem.)	.863	-	-	.830
3 VVD (liberals)	.799	-	-	.733
4 health care	.825	-	-	.499
5 post delivery	.760	-	-	.513
6 public transport	.820	-	-	.614
7 police	-	.605	.736	-
8 social security	-	.547	.698	-
9 road system	-	.550	.731	-
10 income	-	.713	.791	-
11 housing	-	.741	.762	-
12 gen. living situation	-	<u>.603</u>	<u>.692</u>	-
<i>mean correlation:</i>	.820	.627	.735	.669

MEMORY EFFECTS IN MTMM STUDIES

This table shows very clearly that, except for the first three items, the correlations between two measures of the same variable after 9 minutes is much higher than after two weeks. This effect is also clear for the mean correlations. We can see from the means that the items which had higher correlations after 9 minutes have lower correlations after 2 weeks. Although the topic might have some effect (for the first three items) these results definitely suggest that the memory effect is very strong and it seems most likely that all estimates will be biased upwards.

In table 2 the estimates of the MTMM models for the same items have been given. The solid line in the table indicates when the two week interval occurred for a specific item. The two week interval occurred for the first 6 topics after the second observation. For items 7-12 the two week interval occurred after the first observation of the variable.

table 2: MTMM parameter estimates for validity and reliability

	validity			reliability		
	10cat	line	5cat	10cat	line	5cat
1 PvdA (labour)	.881	.926	.919	.884	.906	.859
2 CDA (chr. dem.)	.858	.939	.912	.900	.900	.873
3 VVD (liberals)	.855	.904	.837	.892	.844	.765
4 health care	.946	.873	.575	.915	.854	.416
5 post delivery	.895	.852	.598	.820	.817	.469
6 public transport	.920	.889	.685	.866	.893	.563
7 police	.668	.868	.867	.632	.880	.787
8 social security	.657	.820	.855	.596	.790	.765
9 road system	.652	.810	.888	.607	.813	.813
10 income	.744	.895	.900	.719	.869	.835
11 housing	.808	.891	.856	.813	.877	.770
12 gen. living situation	.691	.844	.816	.684	.790	.699

l = a two week interval between observations

This table indicates that the instruments which were employed together in one interview have much higher validity and reliability estimates (except for the first three topics) than the items which have been asked separately, irrespective of whether this observation occurred before or after the interview with the two observa-

table 3: Mean parameter estimates for validity and reliability (derived from table 2)

	validity			reliability		
	10cat	line	5cat	10cat	line	5cat
items 1-6:	.893	.898	.754	.880	.869	.658
items 7-12:	.703	.870	.864	.675	.837	.778

l = a two week interval between observations

tions for these variables. In table 3 the means of these coefficients are given and this result is here even clearer. These results suggest that the memory effect leads to an overestimate of the quality measures of the variables which have been observed in the same interview. In this data set we did not find any indication of reduced correlation due to so-called "socratic effects".

Given this conclusion and the size of the effects of the time difference on the estimated coefficients, we will concentrate for the rest of this chapter on the memory effects and try to establish first how many people remember their answers correctly.

MEMORY OF PREVIOUS ANSWERS

In order to determine whether a respondent could remember his or her previously given answers the following question was asked:

"Can you remember exactly what you answered on the question about ?

This question measuring the self indication by the respondents was followed by a question which asked for the exact response. If the respondents claimed that they remembered the answer, we asked:

"What number was given by you?"

If the respondents said that they did not remember the answer the following question was asked:

"Estimate the number which was given by you:"

MEMORY EFFECTS IN MTMM STUDIES

The first question gives us an indication of how many people think that they can correctly remember the answers. The last two questions provide us with an estimate of how many people can exactly reproduce their answers. We are purposely not saying that they remember the answer correctly because it is quite likely that some respondents will give a correct answer by chance, even though they no longer recall the previous answer. In tables 4 and 5 the results of the analysis of these three questions are summarized.

table 4: Self-indication of memory

	memory %	no memory %	no memory, no answer %	
<i>±9 minute interval:</i>				
1 PvdA (labour)	72.0	22.5	5.5	N=793
2 CDA (chr. dem.)	73.6	21.4	4.9	N=793
3 VVD (liberals)	70.9	22.7	6.4	N=793
4 health care	79.6	18.0	2.4	N=744
5 post delivery	84.9	12.6	2.4	N=744
6 public transport	77.9	20.1	2.0	N=744
<i>two week interval:</i>				
7 police	18.0	70.3	11.7	N=710
8 social security	16.8	68.7	14.5	N=710
9 road-system	15.3	66.9	17.7	N=710
10 income	27.8	61.6	10.7	N=666
11 housing	38.9	55.4	5.7	N=666
12 gen. living situation	27.9	63.8	8.3	N=666

We distinguish between the question repeated within the same interview and in two different interviews. Within the same interview the time between the first question and the memory question on the same item was on average 9 minutes. For the first 6 topics the memory question was asked within the same interview. For the last 6 topics the respondent had to go back in memory more than 2 weeks.

In table 4 we see that on average 76% of the respondents claimed to remember exactly what they had said previously in the same interview whereas 24% did not remember it. Of all respondents 4% said that it was impossible to give an estimate of their previous answer on these first six topics.

For the topics 7 to 12 we have comparable results only for the two week interval. Approximately 24% said that they could remember their answer exactly and the remainder said that they

CHAPTER 6

could not remember their answer at all. For both sets of answers these percentages do not have to indicate how many people indeed remember their previous answer correctly. In order to check this assumption we asked further extra questions about the given number. In table 5 the results of these questions are summarized.

table 5: Percentage of correct reproduction of previous answers

		self-indication*	
		memory	no memory
<i>±9 minute interval:</i>			
1	PvdA (labour)	57.1	66.2 N=571 33.9 N=222
2	CDA (chr. dem.)	57.9	66.4 N=584 34.1 N=209
3	VVD (liberals)	59.9	68.5 N=562 39.1 N=231
4	health care	67.0	73.4 N=592 42.1 N=152
5	post delivery	74.1	79.3 N=632 44.6 N=112
6	public transport	72.1	79.1 N=579 47.3 N=165
<i>two week interval:</i>			
7	police	40.6	57.0 N=128 37.1 N=582
8	social security	33.1	42.9 N=119 31.1 N=591
9	road system	32.8	52.3 N=109 29.2 N=601
10	income	41.1	53.5 N=185 36.3 N=475
11	housing	48.7	63.7 N=259 39.2 N=401
12	gen. living situation	47.7	54.8 N=186 44.9 N=474

* all group means are significantly different at a .05 prob. level, two tail T-test

The results in table 5 clearly indicate that not all respondents who claimed to remember their previous answer could reproduce it. Only approximately 70% of this group could reproduce the correct number. However approximately 36% produced the correct answer even though they said that they did not remember their previous answer. It should be noted that for these respondents it did not matter whether the question was asked within the same interview or after two weeks, in the second wave of the panel study. Also, the respondents who after two weeks claimed that they could remember the previous answer did not perform any better on the memory task. In this group the percentage of correct answers was close to 36% as well.

These results suggest that 36% of respondents can produce the same answer to a repeated question even though they have forgotten their previous answer. This phenomenon is not surprising because most respondents are not likely to have changed their opinion in such a short time period. Therefore, there is a relatively

MEMORY EFFECTS IN MTMM STUDIES

high probability producing the same answer. Given our results we estimate the probability that they produce the same answer again without any memory effect at 36%. It follows that of the respondents who can reproduce the answers correctly after 9 minutes approximately (70%-36% or) 34% is probably due to memory effects while the others reproduce the result because of the stability of their opinion.

It would be interesting to know how high this percentage is after 4, 8, 12, 15 or even 20 minutes. If there is a considerable decay in memory after some time, this information would be useful in the design of MTMM studies. This aspect can be studied because a computer-assisted data collection procedure was used in which the length of the interview differed, depending on the branches of the questionnaire the different respondents had to answer: for some the time between when the question was first asked and the memory question was less than 3 minutes while for others the interval was more than 20 minutes. If we split up the sample into different groups according to the time between the question and the memory questions and recalculate the percentages of correct answers for the different topics we get the results presented in table 6, which are graphically represented in figure 2.

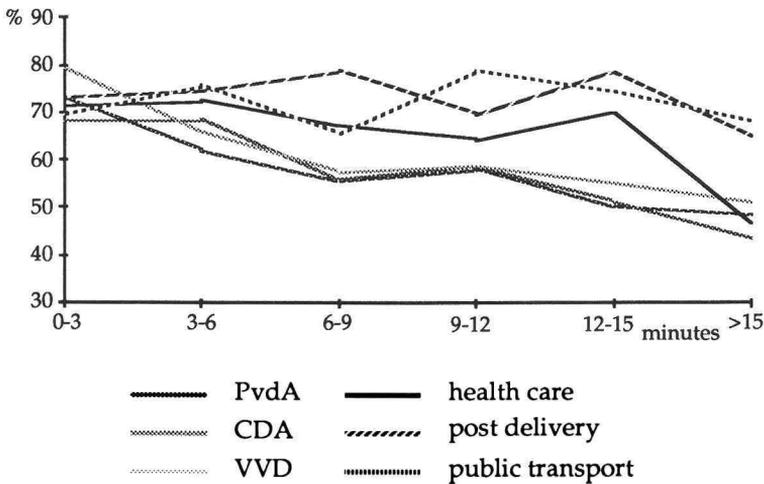


figure 2: The percentage of correct answers for different time intervals

table 6: The percentage correct answers for different time intervals (in minutes)

	0-3	3-6	6-9	9-12	12-15	15-20	> 20
1 PvdA (labour)	73.0	61.9	55.5	58.1	50.0	48.6	46.2
2 CDA (chr. dem.)	68.3	68.5	55.9	58.8	51.2	43.1	43.6
3 VVD (liberals)	79.4	65.8	57.3	58.8	54.9	52.8	46.2
4 health care	71.4	72.7	67.0	64.0	69.9	40.9	52.8
5 post delivery	73.2	74.6	78.8	69.4	78.5	61.4	69.4
6 public transport	69.6	75.8	65.5	78.9	74.2	63.6	72.2

In this table and figure we see that for the political topics 1-3 there is a strong reduction in the number of correct answers to the memory question as the time interval increases from 3 minutes to more than 20 minutes. The reduction in correct answers is around 25%. However for the topics 4 - 7, this did not occur.

These results suggest that the time interval between the repeated questions is not the only factor involved. In order to get a better understanding of the effects of different factors on the percentages of correctly reproduced answers we have made a further analysis which will be reported in the next section.

FACTORS WHICH DETERMINE SHORT TERM MEMORY

We hypothesized that the following factors play an important role in the recall of previous answers:

- the time interval between questions
- the type of questions presented between repetitions
- the extremeness of the response to the question when first presented

The first factor has been discussed above. The second factor concerns the type of questions between the first presentation of a question and its repetition. In our earlier experiments we found that respondents are unable to remember an answer to an earlier question if a long series of similar questions has been asked before its repetition. For example, in a judgement study we asked respondents to evaluate 25 occupations with respect to status (Saris, 1982). After these 25 occupations were evaluated by line production, the same question was asked and the respondents were expressed their opinions by magnitude estimation. In this and similar

experiments it was clear that the respondents did not remember their answer to a specific item due to the large number of similar judgments they were asked to give.

In the present experiment the intervening questions were of a political nature, asking about party identification, probabilities to voting for a party, etc. We believe that it is due to this type of question during the time interval that the political questions show more decay in memory than the other questions.

The third factor is very evident. We hypothesize that it is much easier to remember an answer which is represented by an endpoint of a scale than one on any other point of the scale. Not only are the answers easier to remember but the respondent probably also has an extreme opinion which is likely to be quite salient and central for him or her (Schuman and Presser, 1981; Billiet et al., 1985).

To test these hypotheses we have studied the effect of the first factor, the time interval between the two questions (T), on the percentage of correct answers to the memory questions (C) for four different situations which are determined by the other two factors. According to our hypotheses we expect that:

- there will hardly be any effect due to the time interval if subjects have given an extreme response
- the effects will be stronger for the questions on political issues than for those on the social services.

The dependent variable in the analysis was the percentage of correct answers in the different subgroups, determined by three factors :

- time interval
- political topic or not
- extreme response or not

In this analysis a simple linear regression model was tested using the multiple group analysis of LISREL (Jöreskog and Sörbom, 1988) where the different categories are weighted by the number of people in the subgroup. A model was hypothesized where the regression coefficients are zero for the cases with extreme responses. The results are presented in table 7, which shows that this model fits to the data well. Further restrictions could not be introduced because they led to rejection of the model at the 5% level.

In this table we see that for the two conditions where respondents have given extreme answers the effect of the time intervals is zero. There is no effect at all due to the length of time between when the question was first asked and the memory questions. This result is independent of the topic of the questions. On average the

CHAPTER 6

table 7: The effect of the time interval (T) on the correct reproduction of the previous answer (C) under different conditions.

topic	type of responses to the first question:	
	no extreme values	extreme values
political	$C = 59.0 - .94 T$	$C = 80.6 + 0.0 T$
not political	$C = 75.4 - .50 T$	$C = 71.4 + 0.0 T$

$$\chi^2 = .43 \text{ df} = 2 \text{ prob} = .8$$

respondents correctly reproduce the answer they have given earlier in 75% of the cases. Assuming that in 36% of the cases the correct answer is due to stability of opinion, we concluded that for a large number of respondents (40%) a memory effect can be expected to occur.

Where no extreme answers have been given the effect of the time interval (T) depends on the topic of the repeated questions:

- if the questions between the first question and its memory question are of the same type (political), then every extra minute in the time interval will reduce the percentage of correct answers by approximately 1% (on average).
- if the the first question is of a different type than the intervening questions (non-political) the effect of the time interval is only half as large: an increase of two minutes reduces the number of correct answers by 1%. In addition, the mean percentage of correct answers is also much higher. Respondents thus have a better recall of their previous answer whatever the time interval.

CONCLUSIONS

In this chapter we have shown first of all that the effect due to the interval between the repeated observations on the size of the correlation coefficients, and consequently on the estimates of the MTMM model parameters, can be considerable.

Secondly our analysis has indicated that these increases in correlations and estimates of the parameters are probably due to the large number of respondents (34%) who can reproduce their responses to previous questions after a short period of time. We also found that in approximately 36% of the cases the answers were correctly reproduced even though the respondents said that they could not remember these answers (after a period of two weeks).

These correct answers are not due to a memory effect but to the stability of opinion over time.

Thirdly, it has been shown that the possibility of remembering the previous answers is not only a function of time. Two other factors played an important role: extreme responses to the questions when first presented, and the type of questions asked between their repetitions.

In this study respondents who gave extreme answers were very capable of reproducing their previous answers and this capability was not affected by longer time intervals between the presentation of the questions within one interview and by the type of questions asked in the interval.

These factors did have an effect when respondents did not give extreme scale points to the questions of interest. If we can accept the results obtained and are willing to extrapolate over time and generalize for other topics, which is of course very questionable, then the following practical conclusions could be drawn:

1. In a situation where similar questions are asked between two repeated observations of the same variable, the memory effects will disappear if the time interval between the observations is at least 15 minutes. Approximately 30 to 35 questions can be asked in such a time interval. This result is in agreement with our experience with judgement tasks with more than 25 similar items (Saris, 1982).
2. In a situation where questions of a different type have been asked between the repeated observations a time interval of 80 minutes can eliminate the memory effect completely.

These results pose serious problems for the MTMM approach to evaluate the quality of measurement instruments. It should be clear from the RMM approach suggested by Saris in this volume that any procedure based on repeated observations would also have similar problems. The memory problems are actually even more serious than the discussion here has indicated as the MTMM and RMM approaches require more than 2 observations for each variable.

These problems require further consideration in the design of such evaluation studies. A few remarks can be made immediately:

1. If the study is done in one wave one should use a very long interview with many questions of the same kind concerning the same topic between the repeated observations.
2. It is worth checking whether only two methods, and therefore only two repetitions of the same topic, can be used instead of three. This would reduce the problems considerably.

3. Panel studies using one method per wave and at least two weeks between the waves should be considered as a possible design. It would be necessary however to check for changes in opinion between the waves. This can be done by a quasi experimental design where under one condition exactly the same questions are used at three points in time and the Quasi Simplex model (Wiley and Wiley, 1970) is used for the test of the change in opinion.

The results presented in this chapter suggest that if these recommendations are not taken into account in an MTMM study it is most likely that memory effects will play an important role and that all validity and reliability estimates will be subsequently overestimated. Our guess is that that this has occurred in most MTMM studies to date.

REFERENCES

- Billiet, J., Loosveldt, G., & Waterplas, L. (1984). *Het survey-interview onderzocht. Effecten van het ontwerp en gebruik van vragenlijsten op de kwaliteit van de antwoorden*. Leuven: Sociologisch Instituut.
- Jagodzinski, W., Kühnel, S.M., & Schmidt, P. (1987). Is there a "socratic effect" in nonexperimental panel studies? *Sociological Methods and Research*, 15, 259-302.
- Jöreskog K.G., & Sörbom, D. (1988). *LISREL 7: A guide to the program and applications*. Chicago: SPSS.
- Pijper, W.M. de, & Saris, W.E. (1986). *The formulation of interviews using the program INTERV*. Amsterdam: Sociometric Research Foundation.
- Saris, W.E. (1982). Different questions, different variables. In C. Fornell (Ed.), *A second generation of multivariate analysis: Vol. 2. Measurement and evaluation*. New York: Praeger.
- Saris W.E., & Putte, B. van den (1988). True score or factor models: a secondary analysis of the Allbus test-retest data. *Sociological Methods and Research*, 17, 123-157.
- Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys*. London: Academic press.
- Verwey, M., Saris, W.E., Mosselman, K., & van Doorn, L. (1989). Tele-interviewing in practice: Household incomes show large monthly fluctuations. *Marketing and Research Today*, 17, 230-240.
- Wiley, D.E., & Wiley, J.A. (1970). The estimation of measurement error in panel data. *American Sociological Review*, 35, 112-117.