

The Time-line as a Device to Enhance Recall in Standardized Research Interviews: A Split Ballot Study

Wander van der Vaart¹

A split ballot experiment was performed on a time-line procedure that was designed to increase the accuracy of responses to retrospective questions. The time-line unites several aided recall properties and was applied to help respondents to reconstruct their educational history. The data were collected during two main waves (1987 and 1991) of a longitudinal social survey in the Netherlands ($N = 1,257$). In both 1987 and 1991 respondents were asked about their educational history from August 1983 on. The agreement between the 1991 reports and 1987 reports about the period 1983-1987 was used as a measure of recall accuracy. It was hypothesized that the time-line would enhance recall accuracy regarding the number of educational courses attended, the starting year of the courses, and the entire set of types of courses attended. Additionally, it was expected that the time-line would be especially helpful if the difficulty of the recall task was high – that is to say, in the case of a high frequency or low saliency of the courses followed. The general picture of the results is that the time-line procedure improved data quality in most conditions and never resulted in inferior data quality, supporting the assumption that it indeed may enhance recall.

Key words: Data collection; recall accuracy; memory cues; life history.

1. Introduction

Although social careers and life histories of individuals are often reconstructed by using data gathered with retrospective questions, researchers generally agree that serious questions can be raised about the accuracy of retrospective data. Notwithstanding the fact that information may be intentionally omitted or distorted, methods research has shown that retrospective data suffer from recall errors such as misrepresentation, omission, underreporting and telescoping (e.g., Bernard et al. 1984; Eisenhower et al. 1991; Schwarz and Sudman, 1994; Sudman and Bradburn 1973; Van der Vaart 1996).

Since retrospective questions are often unavoidable and at the same time appear to be error prone, it is important to develop questioning procedures that may enhance recall. High demands are made upon the respondent's memory, especially if retrospective questions are used to reconstruct the respondent's educational or professional career or even life history over a long period of time. To help respondents to fulfill this difficult task,

¹ W. van der Vaart, Department of Social Research Methodology, Vrije Universiteit Amsterdam, De Boelelaan 1081-c, 1081 HV Amsterdam, The Netherlands. Email: w.van.der.vaart@fsw.vu.nl

An earlier version of this article was presented at the International Conference on Questionnaire Development, Evaluation and Testing Methods, November 2002, Charleston, SC, U.S.A.

Acknowledgments: This study was part of the research program "The process of social integration of young adults," directed by the Department of Social Research Methodology and the Department of Work and Organizational Psychology at the Vrije Universiteit Amsterdam. The author is indebted to Johannes van der Zouwen and Wil Dijkstra for their share in the establishment of the study.

the present study offers a time-line procedure that may be used as a recall aid, supplementary to a standardized questionnaire. The effect of adding the time-line (or not) was tested in a split ballot study of the educational history of young adults.

The research question of the present study is whether the time-line procedure may enhance recall accuracy in a social survey and to what extent its effects depend on the difficulty of the respondent's memory task. After a review of the potential memory effects of the time-line procedure and its main characteristics, the results of the split ballot experiment on the time-line procedure are presented.

2. Stimulating the Respondent's Memory

Questioning procedures aimed at the prevention or reduction of recall error in surveys may be divided into aided recall procedures and bounding procedures (Sudman and Bradburn 1974). Aided recall refers to procedures that are designed to reduce recall loss by providing the respondent with memory cues, whereas bounding is a specific type of aided recall that refers to procedures that are designed to prevent telescoping – that is, perceiving events as less (or more) remote in time than they actually are. Bounding can be accomplished by clearly demarcating the period the respondent has to report about, for example by employing landmarks – i.e., specific, salient events – or by providing (other) memory cues regarding time periods and dates (Neter and Waksberg 1964; Sudman and Bradburn 1973). As will be illustrated below, the time-line unites properties of both aided recall and bounding in order to stimulate the respondent's memory.

According to Shum and Rips (1999), a fairly consistent picture of memory emerges from the assumptions that contemporary cognitive theories share. They state that two ideas are central to these theories. Firstly, autobiographical memory is memory for representations of personal events. Secondly, people retrieve these representations by describing a sufficient number of their parts. However, the representations may be conceived not as discrete entities but as a distributed pattern of activation across many processing units in memory, such as those formulated by connectionist models of mental representation (Smith 1999). The idea of distributed representations fits with the often-employed assumption that the information about events or states probably does not have a fixed location in memory, but is organized in multiple frames of reference (e.g., Johnson and Hasher 1987; Reiser et al. 1985). That is to say that information can be organized in connection to all sorts of causal factors, circumstances, and consequences. It is often assumed that the organization of autobiographical memory can be represented as a hierarchical network, in which specific types of events are related to more general types of events (Belli 1998; Belli et al. 2001).

The notion of “multiple frames of reference” is of great importance with respect to memory cues. Recall of a certain past event stored in a respondent's memory would require a search through many cognitive structures. This, however, is a hard task to perform. Through the provision of cues, the search through memory may be more directed, and as a result, become easier. But since respondents appear to organize their memory in different ways and even the same person may show little consistency in organizing memories of different types of events (Sudman et al. 1996), providing effective cues is a complicated issue. As formulated by Tourangeau et al. (2000, p. 96): “There are clearly

limits to the extent to which surveys can provide respondents with cues that will be helpful [. . .], since surveys can't tailor questions to individual respondents' encoding."

These limitations apply especially to structured questionnaires as used in large-scale surveys in which aided recall and bounding techniques should provide effective prompts without altering the questionnaire and disturbing the flow of the interview. In this respect, a time-line that is used supplementary to a questionnaire may be a suitable instrument. Furthermore, it fits the presumptions of "hierarchical memory structures" and "distributed representations": it provides multiple and different types of cues and it offers great flexibility when it comes to applying cues to the respondent's situation.

A further advantage of this device is that while aided recall is predominantly directed at the mere recall of events and bounding at the dating of events, the time-line explicitly combines both properties by utilizing landmarks. Landmarks, in particular personal landmarks, help respondents to date events more accurately, presumably by converting absolute time estimates – exact dates – into relative time estimates – the occurrence since the landmark (Baddeley et al. 1978; Loftus and Marburger 1983).

The importance of combining aided recall and bounding techniques is stressed by Auriat (1993), who examined findings that were obtained without recall aids. It appeared that dating errors were correlated in the recall of chronologically structured series of events: an incorrect date given for one event was more likely to lead to an incorrect date for the next event. Similarly, Auriat demonstrates that forgetting an event is correlated with dating errors for following events. The use of techniques like the time-line might diminish correlation between errors.

Procedures like applying the time-line are designed to enhance recall accuracy. However, their effects may not be equal for all respondents but instead depend on the respondent's actual situation in the past (Van der Vaart et al. 1995; Van der Zouwen et al. 1993). For example, the difficulty of giving an accurate answer to a question like "Which types of educational courses did you attend in 1999?" may be comparatively great for respondents who attended many small courses. Respondents who are confronted with a very difficult task may benefit more from recall aids than respondents whose task is easier.

3. Characteristics and Possibilities of Time-line Methodology

3.1. Time-line procedures and their effects

The time-line procedure involves the application of an aided recall device, a graphical time-line, *in addition to* questions from a standardized questionnaire. It generally is directed at one theme or life domain only and can be conceived as a single line or a few parallel lines divided into time-units (see the appendix for an example of the time-line). The respondent is instructed by the interviewer to mark particular events and states on the time-line and indicate (personal) events that can function as cues or landmarks, making it easier to recall and date the information in question. While answering the questions from the questionnaire, the respondent may refer to his or her time-line (Jobe and Mingay 1989; Means et al. 1988). The time-line possesses several aided recall properties. Firstly, by placing different activities within one time frame the respondent is helped to relate, visually and/or mentally, the timing of several kinds of events. Inconsistencies in reports

are more easily discovered and one event may prompt the recall of another. Secondly, detailed sequences of events are easier to report since they can be marked graphically on the time-line.

Although no experimental studies regarding time-lines have yet been carried out, the existing nonexperimental results demonstrate that time-line procedures may have beneficial effects on data quality. Means and colleagues (1988) reported strong time-line effects in their study of the recall of medical visits over the previous twelve months. They obtained the respondents' medical records, which permitted an assessment of the validity of the responses. The respondents were first asked to recall medical visits without any aid. After that, one or two measures were taken to improve recall, which included cueing and using additional questions. Finally, they let respondents write down some personal landmarks on a time-line and found that its application brought about a considerable increase (18%) in the number of medical visits reported. This result was not obtained at the expense of a higher "false alarm" rate (i.e., reporting nonexistent visits), which meant that the time-line improved recall considerably. Similar nonexperimental outcomes were obtained by Jobe and Mingay (1989). On an eighteen-month time-line, respondents marked important personal events that could be accurately dated. After this, the health interventions (such as medical tests) they had experienced were written on the time-line. Applying the time-line improved recall from 32 to 60 percent of these interventions.

3.2. Time-line-related methods

Additional support for time-line procedures may be derived from empirical findings regarding event history calendars, since the two methods have many aided recall properties in common. The event history calendar (EHC) is a data collection device that is applied by the interviewer to administer and record the answers of respondents about (sequences of) past events within multiple life domains. Usually, an EHC consists of a large grid in which one dimension represents the time units and the other dimension specifies the life domains and issues regarding which information is to be recorded (Freedman et al. 1988; Tagg 1985). Previous research has indicated that employing an EHC may indeed be helpful. Freedman and colleagues (1988) report a number of nonexperimental EHC studies that found relatively high correspondence between retrospective calendar data and matching actual responses obtained beforehand. Yet since there is no control condition it is not known whether these results are due to the calendar procedure. Belli et al. (2001) performed a field experiment comparing an EHC and a regular questionnaire. The quality of retrospective reports obtained in 1998 on social and economic behaviors in 1996 was assessed using the matching responses from the same respondents collected one year before, in 1997. Their finding that the EHC improved data quality on several topics supports the assumption that its properties do improve recall accuracy.

3.3. Advantages of time-line procedures

Aiming to construct a time-line device that enhances the quality of life history data, it is useful to draw some further comparisons between time-line procedures and EHC procedures. Generally speaking, whether a time-line or an EHC is better suited, depends

first and foremost on the research problem at hand. In addition, each procedure offers different opportunities to aid recall. Four main considerations may be distinguished.

3.3.1. Specific or general life domains

The time-line is directed at one specific theme or life domain only and can be adapted specifically for that purpose whereas the EHC procedure may take advantage of the fact that respondents are asked about a range of life domains that may prompt the recall of another. For the purpose of offering more specific memory cues or cues that stem from outside the target life domain(s), a time-line may be more suitable than an extensive EHC.

3.3.2. Central role of interviewee or interviewer

Secondly, the generation of specific cues is stimulated by the active role of the respondent in the time-line procedure. By marking and drawing information on it, the respondent may personalize the time-line to a high degree. In line with the connectionist models of mental representation (Smith 1999), a great variety of specific, personal cues would be very appropriate to enhance recall. The comment must be added, though, that cues and landmarks spontaneously named by respondents are not necessarily appropriate and sometimes may put them on the wrong track (Campanelli and Thomas 1994). The EHC procedure grants a very central and active role to the interviewer in enhancing recall. It uses a flexible interviewing approach in which there is no strict adherence to asking scripted questions. The interviewer is crosschecking information, actively presenting cues, applying checklists and probing for information (see Belli et al. 2001). The active role of the interviewer is aimed at eliminating internal inconsistencies and enhancing recall, which may be very effective but can also in itself distort accounts (Tagg 1985). The interviewer may be steering too much and the respondent may find the negotiations offensive.

3.3.3. Easy to administer versus complicated interviewer task

A third consideration relates to the difficulty of the interviewer task. Since an EHC is not only meant to stimulate recall but also to administer the responses, it demands complex skills of the interviewer. Moreover, an EHC generally involves a large and complicated instrument that may easily result in coding errors made by the interviewer. Therefore, the use of an EHC often requires intensive training of the interviewers. Freedman et al. (1988) reported that in their case the training time was tripled. The time-line, being a less inclusive instrument than an EHC, is less complicated to handle for interviewers.

3.3.4. Supplementary device versus vehicle to collect and administer data

In contrast to the EHC, the time-line consists of a separate aided recall device that is applied supplementary to a standardized questionnaire. But in spite of its simplicity, it still manages to include similar aided recall properties. The time-line can easily be added to a regular survey questionnaire without affecting the formulation of the target questions. This often is a requirement if longitudinal studies are concerned.

4. Hypotheses Regarding the Time-line Procedure

In the present study a time-line was designed and applied in a field experiment concerning the reconstruction of the educational history of young adults. As stated earlier, it was expected that the time-line procedure would be especially helpful for respondents who were confronted with a comparatively difficult recall task. Therefore two hypotheses were formulated. Whilst the first hypothesis refers to the main effect of the time-line, the second hypothesis concerns the interaction with task difficulty.

4.1. The time-line hypothesis

As compared to the regular questioning procedure, recalling one's educational history with the aid of a time-line will result in more accurate reports with respect to: (a) the number, (b) the dates and (c) the types of educational courses attended.

4.2. The task difficulty hypothesis

The supposed positive effect of the time-line procedure on recalling one's educational history will be relatively large if: (a) the frequency of educational courses attended is high, and (b) the educational courses are little salient to the respondent.

5. Research Design

5.1. Data and split ballot experiment

The data used in this study were collected during two main waves of a national, longitudinal study, which inquired into the process of social integration of young adults in the Netherlands. The first wave took place in the autumn/winter of 1987 and the second in the autumn/winter of 1991. In 1991 the sample contained 1,257 young adults between 22 and 30 years of age. During both waves data collection was performed by means of a standardized face-to-face interview. The questionnaire mainly concerned the respondent's life history with respect to education, work, living arrangements, partner relationships, parenthood, and social life.

The time-line procedure was applied to the respondents' educational history and tested in a split ballot experiment during the second wave (1991). All respondents were asked to report in chronological order all education received since August 1983 up to the 1991 interview; half of them were instructed to complete the time-line before the regular questioning procedure started. The two samples were made comparable by distinguishing respondent groups that were defined by sex, age cohort and interviewer (who was connected with a particular region of the country): in turn the two versions of the questionnaire were randomly assigned within each group. Additionally, during the data analyses it was verified that effects of sex, age and educational attainment did not bias the results. The split ballot results pertain only to respondents who had attended at least one training course during the whole period 1983–1991 ($N = 1,083$); correspondence courses and in-service training courses of companies were not considered.

The national study provided for a standard interviewer training of (at a maximum) three days in which the time-line obtained only limited special attention. The interviewers were

not informed about its expected effects and it was just explained to them how to handle this instrument (see below), just as it was explained to them how to handle show cards, checklists etc. The research staff closely supervised the fieldwork and there were no indications that the use of the time-line overburdened interviewers or respondents (for further accounts of the data collection, see Dijkstra 1993).

In order to estimate the effects of the time-line, the agreement between the 1991 reports – as obtained with or without a time-line – and the matching 1987 reports about the period 1983–1987 was used as a measure of recall accuracy. Since no validating information was available, it was assumed that in the case of disagreement the latter reports were more accurate than the 1991 ones about the same events since these reports had to bridge an additional four-year interval. It should be noted that this assumption may not necessarily hold and instead the 1987 response may be the (most) invalid response; in that, rather hypothetical, case we would underestimate the quality of the retrospective 1991 data. If available, external information like records that already existed in the past may be a better alternative for estimating recall accuracy, though that information also may be incomplete and have other drawbacks (Sudman and Bradburn 1983). To enhance the comparability of the 1991 responses with the 1987 standard, the questions were worded identically and were placed at the same position in the face-to-face questionnaire, which was rather early in the questionnaire.

5.2. *Operationalizations*

The setting of a longitudinal social study imposed inevitable restrictions on the operationalization of the present study. Within the existing study of social integration of young adults the time-line was developed, topics were chosen and measures for recall accuracy and task difficulty were established.

5.2.1. The time-line procedure

The time-line procedure was designed to utilize several aided recall and bounding techniques, starting from two approaches. First, it was intended to provide one time frame for different activities that may stimulate the recall of others. To that end a visual aid was employed, consisting of a grid divided into years and months, covering the period 1983–1991. The topic of interest being education, school years were used instead of calendar years. Each school year encompassed five parallel lines on which different types of information could be written (see the appendix for a filled-in example of the time-line). Second, the method was intended to assign an active role to the respondent and prevent detrimental steering behavior on the part of the interviewer. It involved a paper and pencil procedure that required respondents to mark and write events and states on the time-line themselves. In this way the procedure was to personalize the time-line and enhance the activation of personal cues.

Given this basic approach the time-line procedure was standardized by means of some instructions for both interviewers and respondents. At the start, the interviewers explained the time-line to the respondents, making use of a filled-in example. After that, the interviewers instructed the respondents to fill in the time-line in three steps, in order to build up its aided recall function in a uniform manner.

- First, respondents were asked to indicate their age in the top row of each time-line year (within the month of birth). This was designed to cue related information and make them aware of the time period about which they had to report.
- Second, respondents were requested to mark in the bottom row of the time-line important personal landmarks that they could date fairly accurately, such as hospitalization, a wedding, and the birth of a child. These personal landmarks were designed to enhance recall accuracy regarding (in the end) dates of educational courses and might also function as cues for being enrolled or not on (which type of) a course.
- Finally, respondents were asked to fill in simultaneously all educational courses, jobs and periods of unemployment, to be indicated in the remaining rows. While the target information concerned educational courses, these other life domains were designed to provide multiple frames of reference and consequently enhance recall. Additionally, in order to stimulate precision of recall and to support crosschecking of information, respondents had to complete each month of the time-line. They were told to draw lines if activities continued longer than one month and mark an “X” to indicate the beginning or the end of a period, and “O” if a period continued into the next school year.

After the time-line was completed, the regular questioning procedure on educational history started; respondents were invited to check their completed time-line while answering the questions. The answers were recorded in a standard way in the questionnaire. Thus the time-line was employed as a recall aid only and did not affect question wordings.

5.2.2. Research topics

In both 1987 and 1991 the longitudinal study contained questions on the respondent’s educational history that provided usable data on the following topics.

- *The number of educational courses (“numbers”)*: as reported for the period 1983–1987.
- *The date of an educational course (“dates”)*: year in which respondents started their first educational course within the period 1983–1987. The starting date was selected because it was more accessible to a substantially larger number of respondents than the finishing date. The year was examined rather than the starting month since most courses appeared to start in the expected months of August and September.
- *The types of educational courses (“types”)*: the complete set of types – full-time, part-time during the day, part-time in the evening – reported for the period 1983–1987.

5.2.3. Task difficulty factors

With respect to each topic, the difficulty of the respondent’s recall task in 1991 was assessed by means of information on the topic as provided by the respondent in 1987.

- *Frequency (applicable to the recall of “numbers” and “types”)*: the number of educational courses attended in the period 1983–1987.
- *Saliency (applicable to the recall of “dates”)*: estimated by the type of educational course. A full-time educational course was considered as a more salient kind of

course than a part-time day course, which in turn is assumed to be more salient than a part-time evening course.

6. Results

6.1. Number of educational courses

6.1.1. The time-line hypothesis

It appeared that applying the time-line had a substantive effect on the frequency distribution of signed recall error in the reported number of educational courses ($\chi^2 = 29.22$ $df = 8$ $p < 0.01$). As compared with the regular procedure, the time-line procedure results in a higher percentage of accurate reports (80% versus 76%), a lower percentage of underreporting (7% versus 16%) and finally a higher percentage of overreporting (13% versus 8%). These figures are summarized by the mean signed difference scores given in Table 1, illustrating that on average the number of educational courses is underreported in the regular procedure, while it is overreported if the time-line is applied. Additionally, the table shows that the mean unsigned difference scores are very much the same for the two procedures.²

It must be remarked that over-reportings in 1991 (as compared to the 1987 data) may sometimes indicate an increase in recall accuracy instead of a decrease: courses unjustly omitted in 1987 may be reported accurately four years later. Underreportings in 1991, however, probably stand for additional loss of information since the 1987 data are likely to already be biased by underreports. Keeping that in mind, the overreporting of courses in the time-line condition and underreporting of courses in the regular condition might be considered to be in line with the time-line hypothesis (1a).

6.1.2. The task difficulty hypothesis

As put forward by Hypothesis 2a, it was expected that the time-line would be especially helpful when it came to recalling the number of educational courses if respondents did actually attend many educational courses during the recall period. In line with the expectation, Table 2 shows that the interaction is statistically significant for both mean unsigned and mean signed recall error.

However, in so far as the respondents are concerned who in 1987 reported having attended no courses at all, the pattern of interaction deviates from expectation. In 1991 many of these respondents did report educational courses, probably due to forward telescoping, and applying the time-line in fact results in even higher instead of lower levels of overreporting. This of course contributes to the statistical significance of the interaction effect of the time-line with the frequency of courses. Nevertheless, if those respondents were omitted from the analyses, the interaction effect remained statistically significant.

² The *t*-tests and analyses of variance presented in Section 6 imply the assumption that the underlying data on "mean errors" in recall are normally distributed. Correspondingly, the observed difference scores ("recall errors") showed many small differences and only a few large ones. Given the satisfactory number of respondents in the subgroups, the application of the tests seems to be acceptable.

Table 1. Mean unsigned and signed differences (R-T1) regarding the number of educational courses attended between 1983 and 1987, as reported in 1991 (R) and 1987 (T1), with or without applying a time-line

	unsigned diff.	(sd)	signed diff.	(sd)	N
time-line	.24	(.53)	.07	(.58)	548
no time-line	.29	(.59)	-.10	(.65)	535
t-test	$t = 1.38 \quad df = 1,081 \quad p = .17$		$t = 4.46 \quad df = 1,081 \quad p = .000$		

Table 2. Mean unsigned and signed differences (R-T1) regarding the number of educational courses attended between 1983 and 1987, as reported in 1991 (R) and 1987 (T1), with or without applying a time-line, and by the actual frequency of attendance according to the 1987 reports

		frequency of courses '83-'87					interaction effect
		0	1	2	≥ 3	total	
unsigned diff.	time-line	.47	.21	.14	.38	.24	$F = 5.51 \quad df = 3/1,075 \quad p = .001$
	(N)	(73)	(173)	(221)	(81)	(548)	
no time-line		.28	.19	.26	.67	.29	
	(N)	(64)	(180)	(225)	(66)	(535)	
signed diff.	time-line	.47	.14	-.01	-.26	.07	$F = 2.48 \quad df = 3/1,075 \quad p = .06$
	(N)	(73)	(173)	(221)	(81)	(548)	
no time-line		.28	.00	-.12	-.67	-.10	
	(N)	(64)	(180)	(225)	(66)	(535)	

Corresponding to the task difficulty Hypothesis 2a the time-line appeared to be most effective for respondents who attended relatively many educational courses. Furthermore, the beneficial effect of the time-line was probably underestimated due to underreportings in the 1987 data: in 1987 omitted courses that were recalled four years later with the aid of the time-line contribute to the level of over-reporting.

6.2. Dates of educational courses

6.2.1. The time-line hypothesis

Contrary to the assertion formulated in Hypothesis 1b, the time-line did not help respondents to recall more accurately the year in which the educational course started. A comparison of the frequency distributions of signed recall error indicated that for the two questioning procedures this distribution is very similar ($\chi^2 = 8.89 \quad df = 6 \quad p = .18$). This implies that the time-line and the regular procedure did not differ with respect to forward or backward telescoping of educational courses. On average the time-line yielded a slight reduction in both the signed and unsigned differences between the 1991 and 1987 reports about the starting year, but they were far from being statistically significant.

If recall error was expressed in months instead of years, the frequency distribution of recall error showed a heaping: respondents misestimated the dates mainly in multiples of twelve months. The time-line appeared to have no effect on this heaping phenomenon either.

Perhaps reporting the starting year of the educational course is relatively easy for the respondents, of whom all are comparatively young, leaving little room for the time-line to

Table 3. Mean unsigned differences (R-T1) regarding the starting year of the first educational course begun between 1983 and 1987, as reported in 1991 (R) and 1987 (T1), with or without applying the time-line, and by the saliency of the target educational course

		High saliency ----- Low saliency			
		full-time	part-time day	part-time evening	total
unsigned diff.	time-line	.32	.33	.44	.34
	(N)	(236)	(56)	(49)	(341)
	no time-line	.29	.29	.80	.36
	(N)	(223)	(51)	(45)	(319)
	(N total)	(459)	(107)	(94)	(660)
		interaction: $F = 4.54$ $df = 2/654$ $p = .01$			

improve recall accuracy. This explanation is underpinned by the fact that 92% of the educational courses started in August or September, indicating important regular types of schooling. Findings of Kurbat et al. (1998) confirm that the start and end of regular school terms may be important periods: students remembered more autobiographical incidents from these periods than from any other.

6.2.2. The task difficulty hypothesis

The suggestion that the main effect of the time-line was inhibited by a relatively easy recall task is supported by the results on task difficulty. In line with Hypothesis 2b the effect of the time-line appears to depend on the difficulty of the recall task as indicated by the saliency of the educational course concerned. Table 3 shows that although the time-line did not enhance recall accuracy for both the more salient types of courses, it did have a pronounced effect on the little salient part-time evening ones.³ The relatively low task difficulty for the majority of the respondents (86%) recalling the starting year of a daytime training course clearly inhibited the main effect of the time-line.

6.3. Types of educational courses

6.3.1. The time-line hypothesis

Recall accuracy regarding “types of courses” was established regarding the complete set of all educational courses attended during the recall period 1983–1987 (containing at least one course) and expressed dichotomously as equal or unequal. The 1991 and 1987 reports regarding the 1983–1987 sets were defined as equal if they agreed with respect to all types of educational courses within the set (full-time, part-time during the day, and part-time in the evening). Otherwise the reports were defined as unequal. In order to disentangle recall effects regarding the number of courses from recall effects regarding the types, only those sets were examined for which both the 1991 and 1987 reports agreed on the number of constituting educational courses.

³ Since there appeared to be no relationship between saliency and signed recall error regarding dates, only unsigned recall errors were analysed.

Table 4. Agreement between 1991 reports and 1987 reports concerning sets of educational courses attended during the period 1983–1987: percentage of respondents with equal or unequal sets, with or without applying the time-line; only respondents who reported equal numbers of courses in 1991 and 1987

	equal sets	unequal sets	
	%(<i>N</i>)	%(<i>N</i>)	(<i>N</i> total)
time-line	93.3(363)	6.7(26)	(389)
no time-line	89.6(319)	10.4(37)	(356)
	$\chi^2 = 3.30$ $df = 1$ $p = .07$		

As shown in Table 4, in line with Hypothesis 1c applying the time-line yielded a substantially higher percentage of equal sets than did the regular procedure.

The percentage of equal sets of types within both procedures was this high (Table 4) because only those respondents who recalled an accurate number of educational courses in 1991 were selected. The figures in Table 4, therefore, stem from relatively accurate respondents. Even so, the time-line did improve recall accuracy concerning the types of courses involved.

6.3.2. The task difficulty hypothesis

In addition it was expected that those respondents who found it a relatively difficult task to recall the types of courses would benefit most from the time-line, i.e., those who in 1987 reported a relatively high frequency of courses. However, the results do not support Hypothesis 2a: an interaction between the questioning procedure and frequency is lacking.

Still the time-line did exert a beneficial effect: it reduced the negative relationship between the frequency of educational courses and recall accuracy regarding their types. This relationship was significant in the regular procedure ($\chi^2 = 6.54$ $df = 2$ $p = .04$), but appeared to be neutralized by the time-line ($\chi^2 = 3.17$ $df = 2$ $p = .20$). In other words, the time-line compensated for the more difficult recall task.

7. Conclusion and Discussion

Drawing on existing aided recall techniques and the related social-cognitive ideas, a time-line procedure was developed and applied in a split ballot design, which examined the reconstruction of respondents' educational careers. The results confirm that such a time-line may indeed enhance the quality of retrospective life history data. As Table 5 illustrates, for two out of three topics (numbers and types of educational courses), the time-line enhanced recall accuracy. In addition, for two out of three topics (numbers and starting years), it appeared to be especially helpful for respondents who were confronted with a comparatively difficult recall task.

It is conceivable that the beneficial effects of the time-line are partly due to motivational factors, such as the suggestion of importance and precision to the respondents (Cannell et al. 1977; Sudman and Bradburn 1983). Yet, the finding that the time-line was more effective for respondents who had a difficult recall task (regarding "numbers" and "dates") than for those who had an easier recall task, suggests that it does provide effective recall prompts. In addition, it appeared that adding the time-line to the questioning procedure did

Table 5. Summary of the split ballot results. Hypothesized positive effects of the time-line procedure on recall accuracy as confirmed (+), not confirmed (0) or negative (-)

Time-line vs regular procedure	Recall accuracy		
	Number of educational courses	Starting year of educational course	Type of education
Hypothesis 1: Time-line effect	+	0	+
Hypothesis 2: Interaction with task difficulty	+	+	0

not worsen data quality. This indicates that overall the time-line did not lead to unwanted steering behavior by interviewers and that it had no detrimental effects on respondent burden, but rather alleviated the recall task.

Recent results by Belli and colleagues (Belli et al. 2002) seem to support the statement that time-line-like methodologies affect memory processing. Using verbal behavioural coding they found that their event history calendar enhanced retrieval strategies of respondents as was intended and that these strategies were positively related to data quality. When it comes to the present study it is hard to guess which part of the time-line procedure exerted most effects, the multiple frames of references, the landmarks, the active role of the respondent, the possibility of personalising the time-line, or – as a shower of shot – an overall, distributed activation of memory units. More research is needed into the factors that make the time-line work.

The present field experiment confirmed the outcomes of earlier nonexperimental studies (Jobe and Mingay 1989; Means et al. 1988) regarding beneficial effects of time-lines on the reported “numbers” of events. Furthermore, it added the finding that a time-line also may enhance recall of properties (types of courses) and dates (starting years of courses). Nevertheless, there are two reasons to presume that the time-line procedure has the potential to accomplish larger increases in recall accuracy than this study could reveal.

Firstly, the present study was characterized by rather high levels of recall accuracy, likely due to the fact that all respondents were comparatively young and most of their reports concerned regular daytime educational courses. The main effects on “numbers” and “types” signified that the time-line improved recall accuracy despite the fact that the percentage of agreement between the 1987 and 1991 reports was already relatively high without any recall aid – that is, 76% and 90% for “numbers” and “types,” respectively. Contrary to expectations, the time-line failed to enhance the accuracy of the reported starting year of school terms. Without recall aid, full agreement on the starting year was 71%; and if a one-year interval was applied, the level of agreement reached 95%. It may be that the very restricted range of inaccuracy regarding date recall accounts for the fact that the time-line did not produce significant effects here.

Secondly, the fact that the time-line exerted less effect on the reports of the starting year (no time-line effect) and the types of educational courses (no task difficulty effect), may be partly due to the matching procedure that was applied to ensure valid comparisons between the 1991 and 1987 reports. Both reports of an educational course were matched

and compared only if they agreed on two criteria: the level of education (lower general secondary education, preuniversity education, etc.) and the subject of the course (agriculture, law, etc.). As a result, the more accurate reports were the only ones that remained in the analysis; forgotten or omitted courses were not applicable anyhow. For these reasons, recall accuracy was biased in a positive direction, leaving less room for the time-line to increase recall accuracy.

Some characteristics of the present study likely also confine the representativeness and consequences of the findings. First, the sample consisted of young people only and a generalization of the results to older people must be made with care. Survey research findings illustrate that age is often related to recall accuracy, though not always negatively (e.g., Herzog and Dielman 1985).

Second, to some extent the finding is also limited to the educational domain. There is no saying whether similar effects of the time-line will be found in other life domains. There is also likely an interaction between the age of the respondents and the topic of study. Many of the young people had either recently finished their educational careers or were enrolled at the time of the interview, which would enhance recall.

Third, employing the time-line probably has lengthened the interview time, although this was not clearly identified. The two versions of the split ballot questionnaire differed with respect to several measures and only the duration of the total interview time was established (i.e., about two hours). The fact that the interview lasted on average twelve minutes longer for one version of the questionnaire is most likely due to the time-line included in this version. On the other hand neither the fieldwork nor the results on data quality gave indications of fatigue effects or other negative consequences of applying the time-line.

Still, the procedure may be improved by adding instructions regarding the pace of filling in the time-line. Furthermore, the time-line procedure could gain in efficiency to a great extent when the multiple domains that are filled-in – in the present study jobs and unemployment next to education – are all part of the subsequent questioning procedure and not solely used as memory aids. In any case, one should be mindful of negative side-effects when designing and applying aided recall procedures like the time-line. In so far as an aided recall procedure involves inevitable (organizational) costs, it may be desirable to specify the target group of people who actually need support (e.g., the elderly) and apply the device to them only.

Keeping the restrictions of this study in mind, an answer may be formulated to the research question as formulated in the first section: whether a time-line procedure may enhance recall accuracy in a social survey and to what extent this depends on the difficulty of the respondent's memory task. The overall picture presented by the analyses is that the time-line had mainly positive effects on recall accuracy; and in the case of more difficult recall tasks, these effects were generally stronger. These results show that the respondent's situation in the past regarding the content of the retrospective questions is related to recall error. Moreover, aided recall techniques like the time-line may compensate for these interpersonal differences in recall accuracy. The present study was carried out not within a laboratory setting but within the context of a "real-life survey." This often entails restrictions, because the methodological interests are generally subordinate to the substantive survey. The fact that the time-line appeared to enhance recall accuracy under

these circumstances not only sustains the generalizability of the findings, but also seems promising for further applications in social surveys.

Appendix

A truncated example of a filled-in time-line employed to reconstruct the respondent’s educational history

1984					1985					Age = 24	
Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	March	Apr.	May	June	July
x-----science---X											
					X-----sociology-----X						
					X-car dealer--O						
accident			wedding			party			Spain		
Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	March	Apr.	May	June	July
1985					1986					Age = 25	
Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	March	Apr.	May	June	July
X-sociology--					X---didactics-----X						
O- car dealer---X					Xunem pl.oyed X---teacher-----O						
			L.A.			Dory			Mexico		
Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	March	Apr.	May	June	July

8. References

Auriat, N. (1993). My Wife Knows Best: A Comparison of Event Dating Accuracy Between the Wife, the Husband, the Couple, and the Belgium Population Register. *Public Opinion Quarterly*, 57, 165–190.

Baddeley, A.D., Lewis, V., and Nimmo-Smith, I. (1978). When Did You Last. . .? In *Practical Aspects of Memory*, M.M. Gruneberg, P.E. Morris, and R.N. Sykes (eds). New York: Academic Press.

Belli, R.F. (1998). The Structure of Autobiographical Memory and the Event History Calendar: Potential Improvements in the Quality of Retrospective Reports in Surveys. *Memory*, 6, 383–406.

Belli, R.F., Shay, W.L., and Stafford, F.P. (2001). Event History Calendars and Question List Surveys: A Direct Comparison of Interviewing Methods. *Public Opinion Quarterly*, 65, 45–74.

Belli, R.F., Lee, E.H., Stafford, F.P., and Van Heest, K. (2002). Calendar Survey Methods: Association Between Verbal Behaviors and Data Quality. Paper presented at the International Conference on Questionnaire Development, Testing, Evaluation Methods, Charleston, SC, U.S.A.

- Bernard, H.R., Killworth, P., Kronenfeld, D., and Sailer, L. (1984). The Problem of Informant Accuracy: The Validity of Retrospective Data. *Annual Review of Anthropology*, 13, 495–517.
- Campanelli, P. and Thomas, R. (1994). Practical Issues in Collecting Lifetime Work Histories in Surveys. *Bulletin de Méthodologie Sociologique*, 42, 114–136.
- Cannell, C.F., Marquis, K.H., and Laurent, A. (1977). A Summary of Research Studies of Interviewing Methodology, 1959–1970. *Vital and Health Statistics, Series 2*, 69. Rockville, MD: U.S. Department of Health, Education, and Welfare.
- W. Dijkstra (ed.). (1993). *Het proces van sociale integratie van jong-volwassenen: De gegevensverzameling voor de tussenmeting en de tweede hoofdmeting. (The process of social integration of young adults: The data-collection of the in-between wave and the second main wave)*. Amsterdam: VU Uitgeverij. [In Dutch]
- Eisenhower, D., Mathiowetz, N.A., and Morganstein, D. (1991). Recall Error: Sources and Bias Reduction Techniques. In *Measurement Errors in Surveys*, P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman (eds). New York: John Wiley.
- Freedman, D., Thornton, A., Camburn, D., Alwin, D., and Young-DeMarco, L. (1988). The Life History Calendar: A Technique for Collecting Retrospective Data. In *Sociological Methodology*, C.C. Clogg (ed.). Vol. 18. San Francisco: Jossey-Bass.
- Herzog, A.R. and Dielman, L. (1985). Age Differences in Response Accuracy for Factual Survey Questions. *Journal of Gerontology*, 40, 350–357.
- Jobe, J.B. and Mingay, D.J. (1989). Cognitive Research Improves Questionnaires. *American Journal of Public Health*, 79, 1053–1055.
- Johnson, M.K. and Hasher, L. (1987). Human Learning and Memory. *Annual Review of Psychology*, 38, 631–668.
- Kurbat, M.A., Shevell, S.K., and Rips, L.J. (1998). A Year's Memories: The Calendar Effect in Autobiographical Recall. *Memory and Cognition*, 26, 532–552.
- Loftus, E.F. and Marburger, W. (1983). Since the Eruption of Mt. St. Helens, Has Anyone Beaten You Up? Improving the Accuracy of Retrospective Reports with Landmark Events. *Memory and Cognition*, 11, 114–120.
- Means, B., Mingay, D.J., Nigam, A., and Zarrow, M. (1988). A Cognitive Approach to Enhancing Health Survey Reports of Medical Visits. In *Practical Aspects of Memory: Current Research and Issues*, M.M. Gruneberg, P.E. Morris, and R.N. Sykes (eds). Chichester: John Wiley.
- Neter, J. and Waksberg, J. (1964). A Study of Response Errors in Expenditures Data from Household Interviews. *Journal of the American Statistical Association*, 59, 18–55.
- Shum, M.S. and Rips, L.J. (1999). The Respondent's Confession: Autobiographical Memory in the Context of Surveys. In *Cognition and Survey Research*, M.G. Sirken, D.J. Herrmann, S. Schechter, N. Schwarz, J.M. Tanur, and R. Tourangeau (eds). New York: John Wiley.
- Smith, E.R. (1999). New Connectionist Models of Mental Representation: Implication for Survey Research. In *Cognition and Survey Research*, M.G. Sirken, D.J. Herrmann, S. Schechter, N. Schwarz, J.M. Tanur, and R. Tourangeau (eds). New York: John Wiley.

- Reiser, B.J., Black, J.B., and Abelson, R.P. (1985). Knowledge Structures in the Organization and Retrieval of Autobiographical Memories. *Cognitive Psychology*, 17, 89–137.
- N. Schwarz and S. Sudman (eds). (1994). *Autobiographical Memory and the Validity of Retrospective Reports*. New York: Springer-Verlag.
- Sudman, S. and Bradburn, N.M. (1973). Effects of Time and Memory Factors on Response in Surveys. *Journal of the American Statistical Association*, 64, 805–815.
- Sudman, S. and Bradburn, N.M. (1974). *Response Effects in Surveys: A Review and Synthesis*. Chicago: Aldine.
- Sudman, S. and Bradburn, N.M. (1983). *Asking Questions: A Practical Guide to Questionnaire Design*. London: Jossey-Bass.
- Sudman, S., Bradburn, S., and Schwarz, N. (1996). *Thinking About Answers. The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass.
- Tagg, S.K. (1985). Life Story Interviews and Their Interpretation. In *The Research Interview*, M. Brenner, J. Brown, and D. Canter (eds). London: Academic Press.
- Tourangeau, R., Rips, L.J., and Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Van der Vaart, W. (1996). *Inquiring Into the Past: Data Quality of Responses to Retrospective Questions*. PhD-thesis. Amsterdam: Vrije Universiteit.
- Van der Vaart, W., Van der Zouwen, J., and Dijkstra, W. (1995). Retrospective Questions: Data Quality, Task Difficulty, and the Use of a Checklist. *Quality and Quantity*, 29, 299–315.
- Van der Zouwen, J., Dijkstra, W., and Van der Vaart, W. (1993). Effects of Measures Aimed at Increasing the Quality of Recall Data. *Bulletin de Méthodologie Sociologique*, 39, 3–19.

Received February 2003

Revised February 2004