Recording activities with TrueTales: A hybrid calendar-question list-approach for computerized collection of biographical and Time Use data

Maike Reimer[1], Britta Matthes, Ralf Künster

Paper to be presented at the "International Association of Time Use Research Conference"
October 2007
Washington, DC.

Abstract

This contribution will present TrueTales, an innovative data collection approach from the German Life History Study. Like time use data, life history research analyses the sequence, duration, context and timing of activities as well as details about them. Both fields therefore need unbiased reports and face similar challenges to data validity at all stages of date production.

In life history research, calendar or time line methods resembling activity diaries for longer time periods have gained ground; mainly as paper-and-pencil-versions but also in computerized form. They were shown to be superior to traditional standardized question lists with respect to data validity of various temporal information dimensions like number of occurrences, duration, timing and sequence of events. Calendar approaches explicitly acknowledge the cognitive and communicative challenges to validity as integral part of the data construction process and actively address them with tools and methods based on insights into the cognitive and communicative bases of response formation. They support respondent's individual memory cues and reconstruction strategies, they facilitate data edition and are interesting and motivating for respondents and interviewers alike. Issues of standardization and interviewer burden however must be addressed.

In addition to "pure" calendars, hybrid approaches have been developed that combine standard survey questionnaire elements with calendar or time line features. TrueTales is a computerized questionnaire tool that implements the hybrid approach in the collection of standardized retrospective biographical data on time spent in various states like employment, education or unemployment, in numerous residences or in different household and family constellations. Its advantages over a question list approach have been examined in an evaluation study. While it was designed for telephone interviews, principles and technology lend themselves to adaptations for self fulfilling questionnaires, to the recording of other exclusive or non-exclusive states or activities and to time scales of any magnitude.

[1] Correspondence please to: Maike Reimer, Bayerisches Staatsinstitut für Hochschulforschung, Prinzregentenstr. 24, 80538 München, GERMANY, reimer@ihf.bayern.de

1. Time Use Data and Life History Data

Time Use Research and Life History Research share the interest in how human beings spend the time of their lives. Consequently, they also share the challenges to data collection[2] associated with surveying people on their behaviour, its context and especially its temporal features (i.e. time allocation; the sequence, synchronicity, duration and exact timing of events; and changes over time in the form of acceleration, deceleration, increase, decrease and many more) (Stafford, 2006).

Both research areas have come up with similar solutions to these challenges: the collection of data as sequences of temporally extended episodes spent in a finite set of (mutually exclusive or parallel) activities over a defined time span divided in discrete time units and related to a conventional temporal metric (calendar). Such event history data contains an enormous wealth of temporal information on lives and is invaluable for the analysis of societal change (Elder 1985; Mayer & Tuma 1990).

Event History Research typically covers much larger time spans (years instead of days or weeks) and collects multiple timelines instead of defining main and secondary activities. To make this more concrete, let me describe the research project and data work with, the German Life History Study (GLHS)[3]. The study's major focus is on social change in patterns of education and training, labour market entry and processes of family formation (see for example Hillmert & Mayer, 2004; Huinink et al., 1995; Mayer, 1990). Are there changes in the timing, variability, coupling, and sequencing of major life course transitions, such as leaving home, finishing school, starting to work, and starting a family? And to what extent do economic, social, and cultural forces explain these patterns and their changes? The GLHS examines these questions empirically with a series of longitudinal studies of selected birth cohorts from East and West Germany (see figure 1).

Figure 1: The German Life History Study (GLHS)

In all studies, event history data is collected in multiple *domains* - work/employment, primary and secondary education, further education, residence, partnership and household/family - in standardized retrospective interviews. Since all life domains are of equal importance, there is no division in main and secondary activities. The data consists of several sequences of biographical *episodes* per person: time periods spent in one of a finite set of states. I.e. the

[2] They do not however share the Terminology. Therefore, similar ideas are often named differently and vice versa, similar terms might refer to different ideas.

[3] Initially based at the Max Planck Institut fuer Bildungsforschung (MPIfB), Berlin Germany now at the "Centre for Research on Inequalities and the Life Course (CIQLE)" at Yale University

"residence domain" is a sequence of cities lived in, the "employment domain" is a sequence of employment and unemployment phases, and the "family domain" a sequence of household compositions etc. Episodes are separated by transitions (i.e. moves, taking up or losing a job, someone moving into or out of the household), and start and end of each episode are dated.

The GLHS works with interviews, mostly by telephone and supported by computer aided telephone interviewing technology (CATI). In every study, methodological innovations are being developed, introduced and empirically tested. (i.e. Brückner & Mayer, 1998; Hillmert, 2002; Mayer & Brückner, 1989; Wagner & Visser, 2002).

2. The collection of sequences of events: Quality indicators

In standardized surveys, interviewers use standardized instructions, questions and probes to minimize interviewer variance and maximize the probability of getting all the necessary and only the necessary information[4] (i.e., Fowler & Mangione, 1990).

To be useful for sophisticated statistical analysis such as Hazard models, event history analysis or OMA, event histories must be complete (no time unit must be unaccounted for), consistent within and across life domains or between main and secondary activities (free of temporal and substantial contradictions) and above all, valid. Data collected as sequences generally yields more valid estimates than techniques such as Q-lists or generic/stylized time estimates of "normal" days (Belli et al., 2001; Belli et al., 2004; van der Waart, 2004).

Validity is not directly observable[5], but valid reports are also consistent and complete. So gaps and contradictions, in addition to being data flaws in their own right, may be indicators of error and invalid information.

3. Calendars benefits for high data quality

The complete, consistent and valid collection of episode sequences relies on the respondent's cognition and the communication between interviewer and respondent. Data quality can be enhanced by data collection techniques that make this task for both as easy and pleasant as possible.

[4] In Life History research, CATI, CAPI and PAPI are so far predominant. In the following, I therefore concentrate on interview settings and do not consider the self-administered paper or computer questionnaires.

[5] Unless of course there is an external criterion of established truthfulness that the survey reports can be evaluated against, but this is hardly ever the case.

In Life history research, so called event history calendars (also referred to as time lines, life history calendars and many more) have gained ground. These instruments resemble many Time Diaries inasmuch as they

- provide one or more graphic time lines,
- prescribe discrete time units,
- visually depict activities as graphs or strings or lines covering (literally) a definite number of these time units and
- link these to a conventional temporal metric (i.e., years, months, weeks, days).

Methodological research into the cognitive and communicative processes at the foundation of answering a survey question has shown that such tools facilitate and support cognition and communication of both respondents and interviewers and therefore enhance data quality (Belli et al., 2001; Belli et al., 2004; Dijkstra, 1987 - REF!)

a) Cognition

There are several interdependent cognitive tasks involved in answering a survey question which vary depending on what information is sought (i.e., Sudmann et al, 1996; Tourangeau et al., 2000). For the kind of information under consideration here, crucial points are:

- Interpretation of central concepts/words
- Memory access of substantial and temporal information to
    - re-construct episodes and transition
    - re-construct exact dates and times for beginning and end and rounding to the time units given

A calendar can help these tasks to be performed better because they support the strengths and counteract the weaknesses of the cognitive mechanisms at work in performing these tasks. They offer possibilities to personalize instructions, questions and probes, thus allowing respondents to use individual recall strategies. *(ELABORATE)*

b) Communication

Respondents frequently see a survey as a special kind of conversation in which they as individuals are given an opportunity to be heard and to contribute their experience. Calendars provide a common basis for the interviewer and respondents and facilitate understanding thus smoothening the conversation and alleviating the burden of the interview for both (Dijkstra, 1987).

Whilst external incentives play a certain role in the readiness to participate in surveys, intrinsical aspects are stronger and less problematic, especially for the motivation to finish the interview. A smooth conversation is in itself rewarding for the social human being; therefore interviews that can be adapted to the flow of a conversation are motivating for respondents (Schober & Conrad 2002).

Standardized interviews by definition use identical instructions, questions and probes for everyone and therefore can block conversational flow and individual recall strategies. This is especially problematic in heterogeneous samples (for example, in nationally representative samples). Calendars permit flexibility for the interview to function as smooth and conversational as possible (NAMEN).

4. TrueTales: An instrument for the collection of events histories

TrueTales is a filter based CATI-Program designed to support cognition, especially recall, and communication when assessing event histories in standardized survey interviews. It is programmed in Microsoft Access and runs on ordinary PCs. For a detailed description of the program features with screenshots see Reimer & Matthes (2007).

TrueTales is divided in two sections: *data assessment* and after that, *data revision.* During data assessment, all episodes are recorded (with some contextual details) and dated in separate modules - one for each domain. In the data revision section, all information from the separate modules is brought together and displayed graphically against calendar-based time-line[6]. Automatic recursive checks highlight potential errors (gaps and inconsistencies), and interviewer and respondent resolve them together in a dialogue guided by adaptive (filtered) question sequences. In both sections, there are personalized cues and probes and interviewers can in many cases adapt the question order and wording to keep the conversational flow.

5. Is data quality better with TrueTales? An empirical evaluation study

TrueTales cannot help yielding complete sequences – the interview can only be finished when all time units are accounted for. Similarly, each and every inconsistency must be reported back to the respondent and either accepted or revised, before the interview can terminate.

[6] "Pure" calendar methods are being discussed in session xxx. In the project xxx, we use a so called hybrid approach, that incorporates the abovementioned features of calendars / time diaries with normal questions.

But since event sequences can be complete and consistent and still contain errors, we will now report findings from an experimental evaluation study that assesses the data quality obtained from administering TrueTales. We conducted a methodological field experiment with 600 GLHS participants that had been interviewed about there lives in 1998. We recontacted them in 2205 and asked them to report on their lives again, and the interviews covered the period from 1993 to 1998 a second time (see figure 2). For this overlap period of about six years, differences between the two reports can be attributed to memory errors in the second interview and the earlier reports are regarded as a standard against which the later reports are evaluated.

Using a split-ballot design, we conducted 300 interviews with TrueTales and 300 with a CATI procedure that had none of TrueTales' features for interview support. Respondents were randomly assigned to TrueTales and to the standard procedure. Since our focus was the recall for episodes and their start and end dates, only a few detail variables were asked for each episode.

Two indicators of data validity were obtained:

- Number and length of episodes are an indirect data quality indicator, with more and shorter episodes indicating higher validity, since errors tend to lead underreporting of events and transitions (Väisänen, 2006, Reimer & Künster, 2007). If TrueTales counteracts memory errors, the event sequences in the TrueTales-condition therefore should yield more episodes in every module than the standard procedure. This advantage should be especially pronounced for those who have complex biographies with more atypical events, such as women.

- Recall tends to decline over time and later reports become therefore more and more inconsistent with earlier reports. If TrueTales counteracts recall errors, the event sequences in the TrueTales-condition should on an individual level correspond more closely to the reports from the first interview.

6. Results of the evaluation study

Figure 3: Number of episodes reported with the standard condition and TrueTales

Figure 3 shows that indeed, for the overlap period, substantially more episodes were reported with TrueTales than with the standard procedure. This is especially the case for less

conventional episodes such as maternity leave and unemployment, showing that these more peripheral, shorter episodes benefit espeecially from the procedure. For Employment episodes and Other episodes, this difference is statistically significant.

Figure 4: Number of episodes reported with the standard condition and TrueTales by men and women separately

Figure 4 shows that women in particular benefit in the domains of vocational training and employment: the numbers of episodes reported with TrueTales is higher for both genders in comparison with the standard procedure, but this increase is more pronounced for women. Moreover, the difference for women is statistically significant for episodes of employment and vocational training. This indicates that TrueTales specifically works better for women's biographies that include more change and more atypical episodes and therefore are systematically disadvantaged with respect to memory problems.

TrueTales superiority becomes also visible in the direct comparison of individuals' reports in the earlier and the later interview. Table 1 shows the percentage of respondents who reported identical numbers of episodes in the earlier and later interviews. The left part displays the correspondence obtained in the Standard condition, the right part the correspondence obtained with TrueTales. TrueTales for all types of episodes yield higher correspondence levels, indicating better recall. While memory error more often leads to forgetting and underreporting, there are also respondents who respond more episodes in the later interview, but both kinds of inconsistent reports are reduced in TrueTales.

Table 1: Individual correspondence level of episodes (only includes those with at least one episode of a kind in the original 1998 interview).

7. Conclusions and Outlook

TrueTales is interesting for Time Use research: it can be scaled to any time scale and is very practical, since it is ACESS-based and easily allows the recording of details for episodes. While it was designed for telephone interviews, principles and technology lend themselves to adaptations for self-administered questionnaires on a home computer or via the Net. There is however no "Build your own TrueTales"- Programming-Kit available. We welcome everybody who works in a similar direction to share thoughts and results.

One word to the issues of standardisation and the role of interviewer. True Tales burdens interviewers with greater responsibilities and endows them with greater liberties as standardized interviewing principles would allow. TrueTales even demands that interviewers respond to individual biographies and individual recall problems with flexible and personalized cueing and probing - albeit guided and supported by automatized rules and routines. Empirical evidence suggests that this may increase interviewer error but decrease the much stronger error components associated with recall (Means et al 1992; Schober & Conrad, 2002) and interpretation of question wording and response categories (Gobo, 2006).

TrueTales is part of the movement towards collaborative or communicative notions of standardization that emphasize a “standardization of meaning” rather than “standardization of wording” (Suchman & Jordan 1990). The interviewer becomes less of a “stimulus-deliverer” and more of a “response guide” and acts as the researcher’s competent "field agent". The collaboration between interviewers and researchers becomes more egalitarian: the researchers have to make plausible their decisions about questions (and the way to put and answer them) to the interviewers and interviewers’ field experience is treated as a valuable source for improvements in the procedure. This requires intense and continuous training and collaboration, and a small team with little turnover and high motivation. A general upgrading and better recognition of field work in the scientific process should be both cause and consequence of such changes in interviewers’ role.

References

Belli, R. F., Shay, W. L., & 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. & Chou, C-H. (2004). Calendar and question-list survey methods: Association between interviewer behaviors and data quality. Journal of Official Statistics 20(2):185-218.

Brückner, E. & K. U. Mayer (1998). Collecting life history data: Experiences from the German Life History Study. Giele J. Z. & Elder, G. H. (Eds). Methods of life course research: Qualitative and quantitative approaches Thousand Oaks, Sage: 152-181.

Elder, G. H. (Ed.) (1985). Life course dynamics: trajectories and transitions, 1968-1980. Ithaca, Cornell University Press.

Fowler, F. J. & Mangione, T. W. (1990). Standardized survey interviewing: Minimizing interviewer-related error. Newbury Park, Sage.

Gobo, G. (2006). Set them free: Improving data quality by broadening the interviewer's tasks. International Journal of Social research Methodology 9(4): 279 - 301

Hillmert, S. (2002). Edition von Lebensverlaufsdaten: Zur Relevanz einer systematischen Einzelfallbearbeitung bei standardisierten Befragungen. ZUMA-Nachrichten 51(26): 120-140.

Mayer, K. U. & Brückner, E. (Eds.) (1989). Lebensverläufe und Wohlfahrtsentwicklung - Konzeption, Design und Methodik der Erhebung von Lebensverläufen der Geburtsjahrgänge 1929-31, 1939-41, 1949-51, Teil 1-3. Materialien aus der Bildungsforschung. Berlin, Max-Planck-Institut für Bildungsforschung.

Mayer, K. U. & Tuma, N. B. (Eds.) (1990). Event History Analysis in Life Course Research. Madison, Wisconsin, The University of Wisconsin Press.

Means, B., Swan, G. E., Jobe, J.B. & Esposito, J. L. (1992). An alternative approach to obtaining personal history data. Biemer, P., Groves, R., Lyberg, L, Mathiowetz, N & Sudman, S (Eds). Measurement errors in surveys. New York: Wiley

Reimer, M. & Künster, R. (2007). Linking job episodes from retrospective surveys and social security data: specific challenges, feasibility and quality of outcome. Manuscript accepted for publication

Reimer, M. & Matthes, B. (2007). Collecting event histories with TrueTales: Techniques to improve autobiographical recall problems in standardized interviews. Quality and Quantity. International Journal of Methodology 41(5): 711-735. DOI 10.1007/s11135-006-9021-y

Sudman, S., Bradburn, N. M. et al. (1996). Thinking about answers: The application of cognitive processes to survey methodology. San Francisco, Jossey-Bass.

Schober, M. F. & Conrad, F. (1997). Does conversational interviewing reduce survey measurement error? Public Opinion Quarterly, 61: 576 - 602

Schober, M. F. & Conrad, F. (2002). A collaborative view of standardized survey interviews. Maynard, D.W:, Houtkoop-Steenstra, H. van der Zouwen, J. & Schaeffer, N. C. (Eds). Standardization and tacit knowledge. New York: Wiley

Stafford, F. P. (in press). Timeline data collection and analysis: Time diary and event history calendar methods. Belli, R. F., Stafford, F. P., & Alwin, D. (in press). Calendar and time diary methods: Measuring well-being in life course research. Newbury Park, CA: Sage.

Tourangeau, R., Rips, L. J. & Rasinski, K. (2000). The psychology of survey response. New York, Cambridge University Press

Väisäinen, P. (2006). Mean of Episode Length as a quality indicator of Time Use Diaries. Paper presented at the 26th IATUR Annual Conference, Copenhagen, August 16 - 18, 2006

van der Vaart, W. (2004). The time-line as a device to enhance recall in standardized research interviews: A split ballot study. Journal of Official Statistics, 20(2): 301-317.

Wagner, S. & Visser, K. (2002). Arbeitsbericht der Nachrecherche zur Studie 'Ostdeutsche Lebensverläufe im Transformationsprozess'. Goedicke, A. Materialband zur Studie „Ostdeutsche Lebensverläufe im Transformationsprozess“. Berlin, Forschungsbereich Bildung, Arbeit und Gesellschaftliche Entwicklung; Max-Planck-Institut für Bildungsforschung.

Figures and Tables

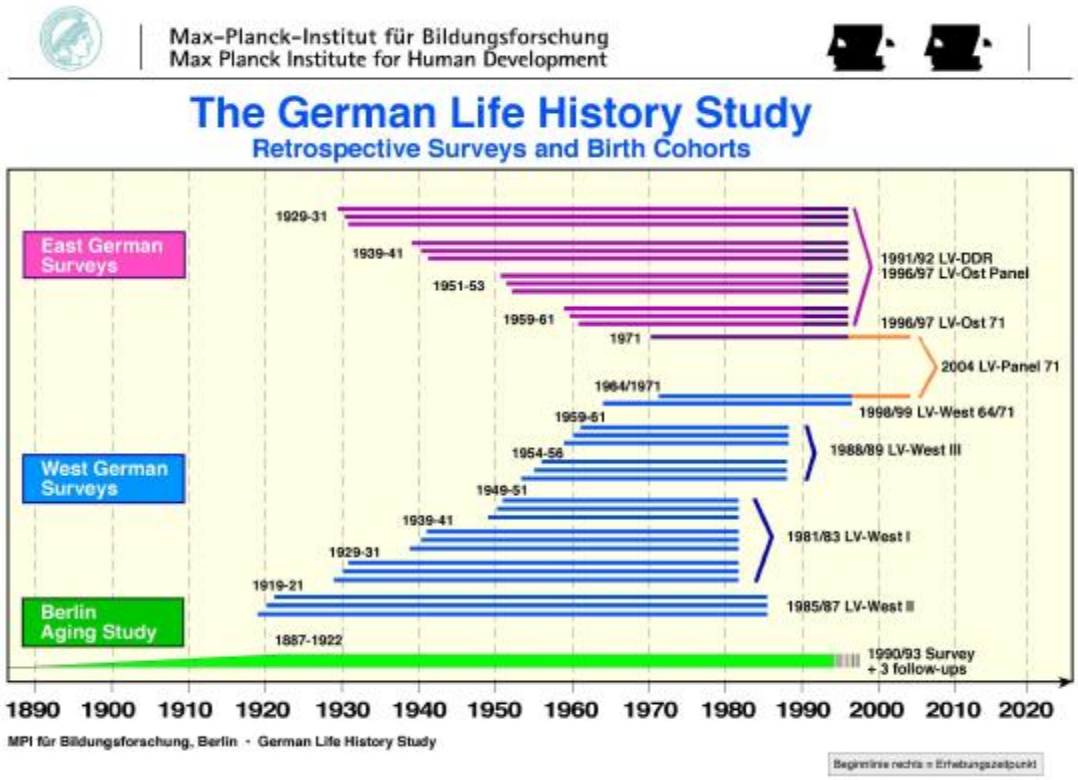

Figure 1: The German Life History Study (GLHS)

Year	2005	2004	2003	2002	2001	2000	1999	1998	1997	1996	1995	1994	1993	1992	1991	1990	1989	1988	1987	...
Age of respondent	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	...
First survey interview									Unemploy-ment1		Job1							Training1		
Second survey interview	Job2			Training2			Unemployment1				Job1									

Figure 2: Design of the evaluation study

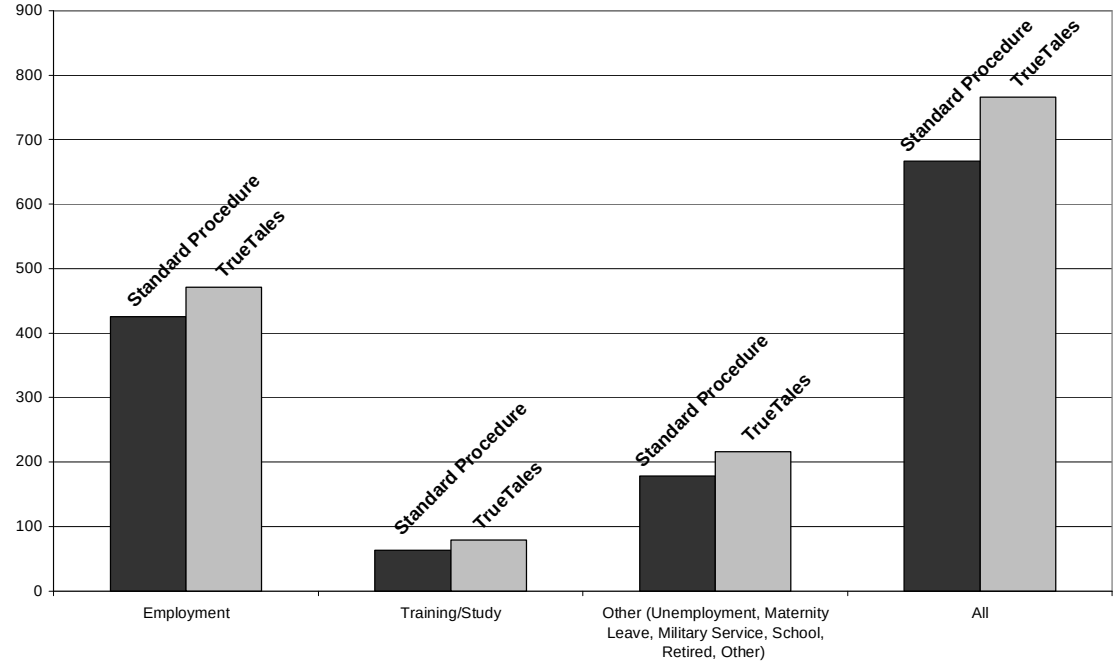

Figure 3: Number of episodes reported with the standard procedure and TrueTales

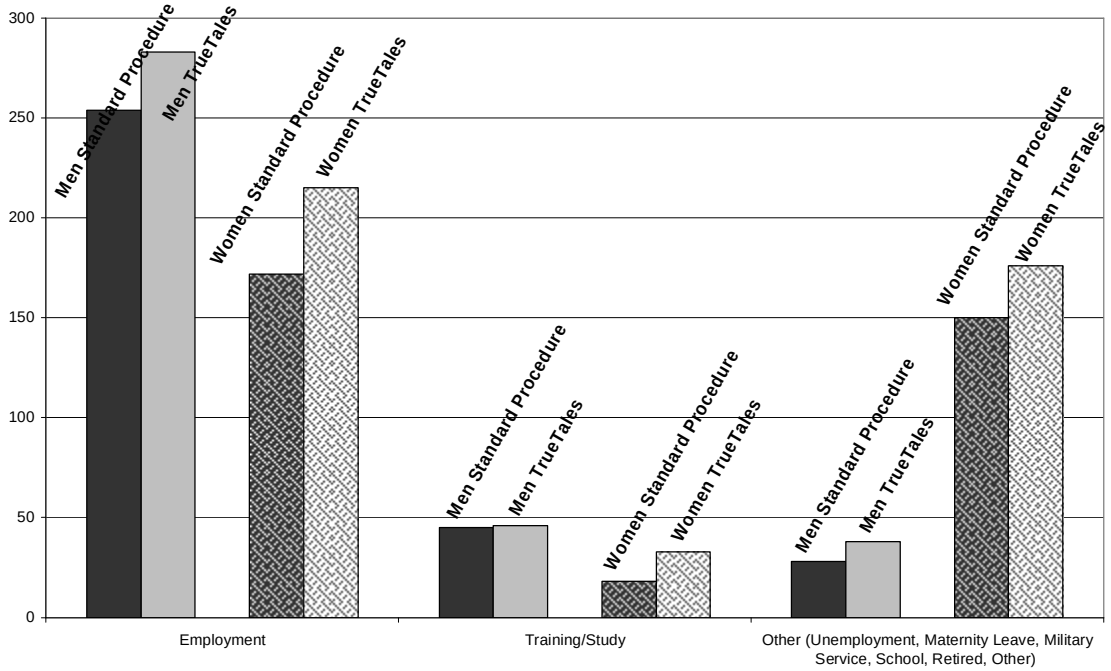

Figure 4: Number of episodes reported with the standard procedure and TrueTales by men and women separately

		Standard Procedure		TrueTales	
		N	*%*	**N**	*%*
Training episodes	same	**23**	*38,3*	**34**	*54,8*
	fewer	**10**	*16,7*	**4**	*6,5*
	more	**27**	*45,0*	**24**	*38,7*
employment episodes	same	**162**	*59,8*	**193**	*68,4*
	fewer	**50**	*18,4*	**40**	*14,2*
	more	**59**	*21,8*	**49**	*17,4*
Other episodes (Unemployment, Maternity Leave, School, Military Service, Other)	same	**49**	*35,8*	**99**	*45,1*
	fewer	**68**	*49,6*	**44**	*20,1*
	more	**20**	*14,6*	**75**	*34,4*

Table 1: Individual correspondence level of episodes (only includes those with at least one episode of a kind in the original 1998 interview).