Materialien aus der Bildungsforschung Nr. 50

Hannah Brückner

SURVEYS DON'T LIE, PEOPLE DO?

An Analysis of Data Quality in a Retrospective Life Course Study

Max-Planck-Institut für Bildungsforschung
Berlin 1995

GW ISSN 0173-3842
ISBN 3-87985-042-9

Materialien aus der Bildungsforschung
In dieser Reihe veröffentlicht das Institut für Bildungsforschung technische Berichte und andere Materialien aus der Forschung, die in der Regel keine abgeschlossenen Forschungsberichte sind, aber dem jeweils interessierten Fachpublikum zugänglich gemacht werden sollen.

Bestellungen werden erbeten an die Verwaltung des Instituts bei gleichzeitiger Überweisung von DM 7,– (einschließlich 7% Mehrwertsteuer) auf das Konto Nr. 0910005885 der Berliner Sparkasse, BLZ 100 500 00.

GW ISSN 0173-3842
ISBN 3-87985-042-9

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Going beyond the traditional social-psychological model of the interview, criticism of the survey method points to the importance of the organizational framework of the interviewing process. Control of interaction in the interview through standardization may lead to an alienation of interviewers and respondents who may react with withdrawal from the task, aggression, or 'output reduction.' Using a large sample of computer-assisted telephone interviews from the German Life History study, I show that a multi-dimensional conceptualization of data quality both in terms of appearances of the data and in terms of their meaningfulness is useful for evaluation. A method for measuring data quality in retrospective life course interviews is presented. Previous findings with respect to respondent effects on data quality, especially of their education, are replicated. Testing hypotheses drawn from organizational behavior perspectives on survey research with respect to the impact of interviewers reveals a complex picture. Specifically, interviewers' experience and productivity are mixed blessings and may, contrary to the hypotheses developed, decrease apparent data quality but enhance meaningfulness.

ZUSAMMENFASSUNG

Theoretische und methodische Entwicklungen in den Sozialwissenschaften führen zur Erhebung von zunehmend komplexen Daten. Anhand der Lebensverlaufsstudie, Geburtsjahrgänge 1954-56 und 1959-61, wird hier versucht, die methodische Forschung zur Datenqualität in Interviews sowohl theoretisch als auch methodisch weiterzuentwickeln. Kritiker der Umfrageforschung betonen die Wichtigkeit der Organisation von Umfragen für das Verhalten von Interviewern und Befragten. Über das sozialpsychologische Modell der Interaktion im Interview hinaus werden Bedingungen der Datenproduktion systematisch in die Analyse der Datenqualität miteinbezogen. Darüberhinaus wird eine Technik aufgezeigt, Datenqualität in computerunterstützten Telefoninterviews zu messen. Dabei wird ein zweidimensionales Modell der Datenqualität entwickelt, welches sowohl Aspekte der Sinnhaftigkeit der Daten im Hinblick auf die Meßkonstrukte (operationalisiert als korrekte Erhebung der Berufswechsel) als auch Aspekte der formalen Konsistenz (gemessen als chronologische Konsistenz der Monats- und Jahresangaben in Berufs- und Wohngeschichte) mit einbezieht. Ergebnisse bezüglich von Befragteneinflüssen auf die Datenqualität aus früheren Studien bestätigen sich, besonders was die Wichtigkeit der Bildung betrifft. Die Analyse von Interviewereffekten auf die Datenqualität ergibt ein komplexes Bild: entgegen der Hypothesen steigt mit steigender Interviewzahl die Sinnhaftigkeit der Daten, während die formale Konsistenz sinkt.

ACKNOWLEDGMENTS

The following text is a slightly revised version of my Masters Thesis at the University of North Carolina at Chapel Hill, Department of Sociology. Barbara Entwisle helped greatly in all aspects of the design, analysis, and writing of the paper. I am grateful to Ken Bollen for support with the data analysis. Furthermore, I thank Peter Bearman for liking the project, and his invaluable help with some thorny theoretical issues. Judith Blau provided other useful tips for interesting readings and generally cheered me up when I got stuck. Various staff members of the Institute for Research in the Social Sciences of the University of North Carolina at Chapel Hill, especially Ken Hardy and Gary Gaddy, helped greatly with data transformation problems and getting my data into Limdep. The Department of Sociology provided computer time, a lot of resources, and a very supportive environment.

In more than one way, this paper was brought into existence by my mother Erika Brückner. Beyond the necessity of a day-to-day survey management routine, she always took the time and energy to teach me an incredible amount of useful skills and knowledge. Many problems discussed in this thesis are a direct off-spring of ideas which evolved in the years I worked under her. I have greatly profited of our inspired if sometimes stormy collaboration. I am also deeply indebted to Karl-Ulrich Mayer for his permission to use the life history data and all the encouragement, resources, and help he made available to me. I would like to thank the editing crew and staff members of the Max Planck Institute for Human Development and Education in Berlin, especially Sigrid Wehner, Jürgen Moka, and Kai-Uwe Schmidt, for help with intractable programming problems, for putting up with my sometimes impatient attitude, and doing all the work. Herrman Lückhoff did a superb job in writing the C-Code for the check program and thus provided a convenient way to measure errors. I am also indebted to the anonymous interviewers whose job commitment and enthusiasm, which made the survey possible, is not sufficiently acknowledged in the remainder of this paper. Richard Hackman first directed my attention towards the literature addressing aspects of organizational behavior in social research. Any remaining errors are, that goes without saying, my own responsibility.

I. INTRODUCTION

Shortly before the 1992 presidential elections, drastically different results of polls regarding the outcome of the election hit the headlines. The New York Times provided a comment by Seymour Martin Lipset (1992), published under the headline: "Polls don't lie. People do." Pre-election polls are an excellent example of the so-called social desirability bias: because voting is socially more acceptable than non-voting, people tell the interviewer that they plan to vote whether they actually do so or not. Furthermore, some people may have never thought about certain questions or topics; upon interviewer's insistence on an answer, respondents might make up an answer, presumably what they think the interviewer wants to hear. Framed differently, a version of Heisenberg's Uncertainty Principle, the extreme case in the natural sciences, is everyday life in social science research. The amount of energy needed to measure an electron's position in the atomic structure changes its position. Asking people questions may induce rather than solicit answers.[1]

Taken at face value, the title of Lipset's comment suggests that survey technology has achieved methodological perfection. Sources of error in the administration of the poll have been eliminated, and dishonest respondents are to blame for remaining bias. To do Lipset justice: According to his comment, differences in predicted voting behavior are reflecting differences in polling practices such as sampling, question sequence and wording, nonresponse follow-up, and contact method.

In this paper, I focus on non-random influences on data quality. Bradburn (1983) distinguishes three sources of variation in data quality of survey interviews: from the characteristics of the task, from interviewers' performance, and from the respondent. Two sources of non-sampling error are considered: error arising from the respondent and from the interviewer.[2] I concentrate on survey data, although some of the questions I address may bear on the nature of empirical evidence in the social sciences more generally.

Furthermore, I am concerned only with factual data, more specifically, with retrospectively collected event history data. The analysis presented in this paper might be especially interesting for sociologists concerned with job mobility and life course research. My indicators of data quality are drawn from variables used also for calculating the duration of events, hence, the dependent variable

---

[1]The same analogy can be found in Bryant (1994).

[2]I cannot treat error through nonresponse here since I do not have the necessary information about non-respondents. Also, I will not analyze error arising from the question since the same questionnaire was administered to all respondents.

in event history analysis. The extent to which mobility and life course disorder (Rindfuss et al. 1987) co-vary with data accuracy of their measures and their predictors bears upon the internal validity of studies using these concepts.

In most research on interviewer effects, the interview is seen as a micro-social system consisting of two roles--interviewer and respondent--linked by the task of transmitting information (Sudman & Bradburn 1974). Advances in survey technology, such as computer-assisted telephone interviewing, allow more control over the situation in which interviews are conducted. At the same time, they may affect data quality in other ways currently not very well understood. ****

1.1. Trends and Problems in Data Collection

The multiple levels on which time enters modern data analyses adds duration, sequence, and speed of events to the classical age-cohort-period model and renders cross-sectional approaches obsolete (Elder 1985, Mayer & Huinink 1990). Life course research has addressed the need for data in a number of ways[3]: by collecting and analyzing narratives (see Bertaux 1981, Bertaux & Kohli 1984), archival data or official records (e.g. Elder & Caspi 1990; Elder & al. 1991; Voges & Zwick 1990), census data (e.g. Oppenheimer, 1974) and, last but not least, by designing social surveys.

Survey design followed one of two strategies: The prospective panel study usually collects retrospective data and present status of respondents in a first wave and than continues data collection following up respondents for some amount of time.[4] Compared to the panel design, the retrospective one-shot study avoids problems of panel mortality and changing measurement conditions (Featherman 1979) but has to face an exacerbated problem of potential recall error.[5] Retrospective designs have been used by the John Hopkins Social Account Project (Blum et al. 1968) and various other studies in the U.S. (e.g. Freedman et al. 1988, Furstenberg et al. 1987). The data used in this paper come from the German Life History Study (Mayer & Brückner, E. 1989; Brückner, E. 1993; Brückner, H. & Mayer 1995).

---

[3]See Elder (1985) for an overview of the history of life course research.

[4]E.g. The Panel Study of Income Dynamics, Duncan & Morgan (1985); German Socio-Economic Panel, Krupp & Hanefeld (1987).

[5]For an entertaining but distressing account of failed retrospective designs see Bernard & al. (1984). A summary of recent research efforts on quality of retrospective data collection can be found in Dex (1991).

Research in sociology and cognitive psychology indicates pitfalls in the retrospective design but also methods to improve recall in surveys. Data quality can be enhanced by constructing the questionnaire along memory structures and retrieval processes.[6] Proximity to 'landmarks' in the memory (Huttenlocher et al. 1988), a complete and detailed reconstruction of life events in chronological sequence (Helling 1987), ordered by domains (such as migration history, education, career history, and family; Niethammer 1989, p. 289,) were shown to support retrieval processes and hence enhance data quality.

As researchers strive to incorporate the complexities of lives into survey design, the role and work of interviewers must be reconsidered. Interviewers have been characterized as "gatekeepers" of the flow of information from respondents to researchers (Cannel et al. 1981). They contact respondents, administer questionnaires and code answers. While factual items may be less vulnerable than attitudinal items to small changes in the stimulus of questions and non-uniform administration, the completeness and accuracy of 'facts' nevertheless depends strongly on interviewers.

Martin (1983) characterizes the social interaction between interviewer and respondent as the most variable and least standardized aspect of data collection. Clarren and Schwartz (1976) remark that "the upper bound of 'crimes' that could be elicited is limited only by the persistence of the interviewers and the patience of the respondent" (p.129). Of course, ambiguous concepts increase variability in reporting as well. What is a crime? What is a job, for that matter? While crystal-clear concepts are desirable, definitions might not always fit into the clear-cut, simple sentences a questionnaire calls for. Even after extensive pretesting researchers cannot always foresee how stimuli are perceived in the field. Hence, the role of the interviewer as a link in the communication between researcher and subject becomes ever more crucial.

Since Rosenthal's (1969) famous study of experimenter bias, researchers have become cautious in carrying out their own experimental designs lest solid deductive work would turn into self-fulfilling prophesy. A similar problem surfaces in survey research: questions should be written in a way that does not suggests answers. After all the work going into contriving non-leading questions, we wouldn't want interviewers to defy our purpose by altering questions or resorting to leading probes in order to get answers. Whenever respondents might ask about the meaning of concepts contained in questions, therefore, interviewers are usually instructed to reply 'whatever it means to you.' Such

---

[6]E.g. Loftus & Marburger (1983), Chassein & Hippler (1987), Rubin & Baddeley (1989), Sudman & Bradburn (1973), Dex (1991). For an overview on the field and implications for questionnaire design see Brückner E. (1990, 1994).

instruction ensures uniformity of stimulus across interviewers and respondents but might entail a price in terms of ambiguity of meaning. Smith (1981) found that GSS respondents had very different things in mind when asked about their confidence in the scientific community. Worse, the referents people had were related to the reported amount of confidence in it.[7] In contrast to research on attitudes, opinions, and emotions, behavioral researchers, trying to explain as exactly as possible what they want to measure, have developed techniques such as aided recall to enhance reporting of behaviors.[8] For data collection of 'facts,' it is advisable to train interviewers extensively in regard to the concepts questions are aimed at (Bradburn 1983). The difficulties which arise when asking about basic facts of life such as marital or labor force status and household composition are amazing.[9] It might therefore be seen as imperative that interviewers understand the concepts and operational definitions of factual items and are able to communicate them to respondents.[10]

On the other hand, interviewer manuals still demand that questions be read exactly as worded and prescribe standardized probes to avoid interviewer-introduced bias. Furthermore, much of interviewing in the survey "industry" is organized as a typical secondary labor market; besides other part-time work such as table waiting, door-to-door selling, and house cleaning, job opportunities in interviewing are found in the classified ads section of newspapers. Interviewing is still hardly a "profession" or an occupational category--rather, it is one of the jobs associated with certain life course transitions, as a (second) job for housewives, students, retirees, and the unemployed. Earnings are rather low, and health care benefits or other characteristics of regular employment are the exception in interviewing. But is interviewing typical of secondary labor market employment in terms

---

[7]A third said they had no referent at all in mind, others thought about particular fields such as medicine, atomic energy, electricity, and the space program. 2% took it to mean their own local community. Apparently, confidence 'in the people running the scientific community' was greater among those who had no referent in mind or thought about the space program (Smith 1981).

[8]While these techniques help avoiding under-reporting, they might on the other hand lead to over-reporting of behaviors due to telescoping effects or social desirability (Bradburn 1983).

[9]For example, how extensive must one's job search be in order to be classified as 'looking for work'? Are in-laws living with the respondent to be classified as relatives or not? Are the children of my boyfriend living in my household to be listed as children or 'other'? What is the marital status of a couple legally separated but still living together in the same household? In my experience, when confronted with questions such as these not even two researchers in the same project agree upon an answer. Much less one can expect interviewers and respondents to do so.

[10]Some of the introductory material to the 'survey craft' suggests almost a convergence between interviewers' and researchers' roles in this respect (Bradburn 1983, Converse & Presser 1986).

of a high turnover and low job commitment? Turnover is certainly high (as many interviewers are hired for one survey and than dismissed, or as they find other, more attractive sources of income). Job commitment is, however, another question. Many interviewers who end up working in the scientific sector of the survey industry are committed workers; they like the communicative aspects of their jobs, they are interested in the studies or feel that they can learn from asking people questions (Converse & Schuman 1974). We might thus find interviewers in a somewhat contradictory occupational role: highly committed workers in a job which is socially and economically constructed as marginal and temporary employment and yet depends arguably on their enthusiasm.

The extent to which interviewers' commitment is beneficial for the quality of survey research might be worth more than a friendly argument between practitioners. I think that it is time to take a sociological approach to the problem of interviewer bias and to study the effects of work place organization and conditions on data quality in order to complement the social-psychological approach to interviewing.

1.2. Criticisms of the Survey Method

The trend towards minimizing interviewer influences by standardizing the interview situation has been criticized by researchers of social interaction. They argue that standardized gates and keepers may preclude the flow of unstandardized information (Pennef 1988, Douglas 1985, Deutscher 1972, Brenner 1978). A second strand of criticism holds that relationships in the triad of respondent-interviewer-researcher, however temporary they may be, bear a remarkable resemblance to those in formal organizations. Respondents and interviewers are placed in situations "similar to those that organizations create for lower level employees" (Argyris 1968, p. 185). Overt or covert withdrawal, low commitment to the task, and conflicting loyalties might result. Argyris (1980) points to these <u>unintended consequences of rigorous research designs</u>.[11] Despite recent trends towards professionaliz-

---

[11]The reader might be reminded of Merton's unanticipated responses found in propaganda analysis during World War II (Merton et al 1956). Researchers could understand their subjects' reactions to movies or other material only by conducting an in-depth, 'focused' interview which uncovered the meaning certain stimuli had for them. Sørensen (1991) translates Merton's concept of <u>depth</u> into what methodologists conventionally address as <u>validity</u>. He suggests that sociologists tend to make progress in establishing relationships between variables, but are less adept in establishing why and how these effects come about. A central problem in survey research is whether meaning can be held constant by holding the question constant. I will address this problem in my paper, however, only in passing.

ation, interviewing is still a typical secondary labor market providing students, housewives, and pensioners with part-time employment. Roth (1978 [1966]) compares "hired hand research" to industrial production, where output reduction and "gold-bricking" are the order of the day.

These mechanisms may not per se inhibit data collection. Interactional trouble and alienation of the actors notwithstanding, researchers may obtain complete and--seemingly--correctly filled in questionnaires. On the other hand, the meaning thus allocated to a question may be inconsistent with the concept the researcher set out to measure. Second, reduced task commitment by both respondents and interviewers may lead to incomplete, superficial, or faked data. As Brenner (1978) puts it, the semantic value of responses is bound by interactional micro-processes which we might ignore. The ironic conclusion common to these critical approaches to social research is that organizational practices aimed at minimizing a known source of bias introduce a new, unknown source of bias.

Recently collected data from the German Life History Study provide a unique opportunity to test this hypothesis. Using a highly standardized computer-assisted telephone interview, retrospective life course data were obtained and subjected to an extensive editing process. Using both the edited and unedited version of the dataset, I develop two measures for data quality: 'visible' and 'invisible' errors. Visible errors are those detected by a computer program designed for consistency checks of event history dates. The identification of invisible errors, namely instances where respondents' answers are inconsistent with measurement concepts targeted with the questions, involves taking into account all available information and is actually similar to qualitative field work.

Research on interviewer effects has concentrated on measuring between-interviewer variability of sample means of single questions or items and on how this variability might be reduced. The approach taken here differs in several respects from previous lines of inquiry. First, rather than looking at single item means, a more comprehensive measure is constructed that combines data related to the entire chronological sequence of the life course into a single indicator of data quality. Second, the in-depth evaluation of micro-processes during data editing allows a quantitative assessment of the seriousness of consistency problems. Interaction coding (Marquis 1969) has contributed a lot to our understanding of social interaction in the interview situation. This technique was less helpful, however, for evaluating the outcome of these micro-processes in terms of data quality. Third, using multivariate techniques, an attempt is made to disentangle response error due to respondents from response error due to interviewers. Finally, I am not aware of a study testing hypotheses derived from organizational behavior theory in the context of a social survey. Previous work has focused on experiments (Rosenthal 1969, Argyris 1968, 1980) and qualitative field work (Roth 1978 [1966]). However, the large majority of studies published in sociological journals utilize survey methods.

I will not be able to compare these data with other data obtained with different procedures. The analysis of effects of respondents' characteristics on data quality, however, addresses the problem of administering a standardized procedure across a heterogeneous sample. The analysis of trends in interviewer performance over time, finally, tests for the development of a 'hired hand mentality' (Roth 1966 [1978]). Concerns about the adequacy of standardized survey instruments have permeated into mainstream methodology, but quantitative empirical work related to this issue is yet sparse (Groves 1987).

II. THEORETICAL AND EMPIRICAL APPROACHES

Generally, from a constructivist perspective, social processes intervene between an account (read: data) and a presumably existing underlying reality. Woolgar (1983) stresses that the constructivist view of science operates under the assumption that accounts (i.e., data) are underdetermined by facts of nature. Two or more alternative accounts can result from the same underlying reality (p. 247) depending on the mediating social interaction. Extending this notion to the survey setting, one might say that obtained sample means may be underdetermined by population means. From this perspective, answers, observations, records--in short, data--reflect outcomes of a complex social interaction rather than reflecting what was the case before the production of the data took place.[12] Based on varying levels of formulations of the problem of data, in the final analysis one can cast doubts on every finding. Without declaring themselves constructivists, survey researchers try to deal with this problem by holding social interaction constant.

Standardizing questionnaires is only one step in this direction. Not only survey questions and their sequence but also probing, feed-back, and other parts of social interaction in the interview are prescribed. Another aspect of the process of standardization is the removal of interviewers' personality and involvement with the respondent from the interview situation. Neutral behavior is preferred over the earlier emphasis on maintaining the 'rapport' with the respondent (Martin 1983, Groves 1989). Similarly, a dispersion of interviews over a large staff of interviewers is preferred over extensive employment of interviewers within one survey.

A great deal of attention has been devoted to the social interaction between respondents and interviewers in the survey. Sources of bias related to the interviewer have been identified. We have come a long way since the classic study of Rice (1929), who found that destitute respondents interviewed by prohibitionists tended to attribute their deprivation to alcohol abuse, while those interviewed by socialists tended to blame industrial society. While Rice interprets the large bias found in his study as an effect of social desirability, Hyman (1954) points to the critical role of interviewers' expectations. He shows that early on in the interview interviewers tend to form an image of respon-

---

[12]Woolgar (1983) remarks that different 'schools' of thought can be classified according to their use of words such as 'in addition too' or 'instead of' in the place of 'rather' in this sentence. This paper is based on the assumption that there is a physical reality which--to varying extent--shows up in scientific and other accounts. The problem I address below concerns the degree to which the flow of events making up a life history is structured according to the 'rules' given by the questionnaire which are only one particular way to structure one's story.

dents' attitude structure. Consequently, answers that do not fit with this image are overlooked or discounted. Furthermore, interviewers tend to evaluate ambiguous responses on the basis of their expectations of what the majority of people would respond. Interviewers' expectations about the ease of asking questions in a particular survey may have significant effects both on overall response rates and response quality on individual items (see, e.g. Sudman et al. 1977, Singer & Kohnke-Aguirre 1979, Singer et al. 1983).

Most research on interviewer effects has focused on attitudinal data.[13] The influence of interviewer characteristics (SES, gender, race, age) and expectations, sometimes in interaction with respondents' characteristics, on reported attitudes as well as nonresponse error gained a lot of attention. Visible and invisible interviewer characteristics are seen as eliciting socially desirable answers from the respondent. Hoag and Allerbeck (1981) tested whether female interviewers, for example, influence respondents towards more positive attitudes towards female labor force participation. While they found no effect on clearly gender-related items, they could show effects on other items such as the acceptance of nuclear power plants and the popularity of certain politicians. Interactions of interviewers' race with respondents' race on item response related to race relations in the U.S. are well documented (e.g. Schuman & Converse 1971, Hatchett & Schuman 1975). Black respondents may express more radical views to black interviewers. White respondents may respond in a more liberal way to black interviewers than to white interviewers. The effects seem to be mediated by respondents' age and social status, and are usually relatively small. Unfortunately, nobody knows how liberal or radical black or white respondents really are, or were.

Research on interviewer effects on behavioral and factual items seems to be somewhat less popular than on attitudinal items (but see Rustemeyer 1977, Hanson & Marks 1958, Freeman & Butler 1976, Bailey et al. 1978). The literature comparing interviewer effects on factual items with those on attitudinal items yielded mixed results: Some studies find larger effects on attitudinal items, some find no difference (see Groves 1989, chapter 8, for an overview). In contrast to attitude questions, however, in some cases the 'true value' (income from tax records, hospital visits from insurance records) can be obtained and compared with the data collected in the survey. Considerable differences are not uncommon (Bernard et al. 1984) but the reasons are not clear. Respondents' inaccuracy, questionnaire design effects, and inadequate interviewer behavior interact in generating

---

[13]E.g. Singer & Kohnke-Aguirre (1979), Singer et al. (1983), Hoag & Allerbeck (1981), Groves & Fultz (1985). See Groves (1989) for a comprehensive overview. Earlier research was reviewed by Sudman & Bradburn (1974) and Weiss (1975.)

inaccurate data, and it is not easy to disentangle these effects. Following Schanz and Schmidt (1984), I propose a multivariate design to separate interviewer effects from respondent effects (holding the instrument effect constant).[14]

2.1. Interviewer Effects and Standardization

Differential interviewing practices yield between-interviewer variability with respect to the distribution of variables in the survey. Depending on interviewer workload, between-interviewer variability inflates the variance of sample means and hence considerably reduces the precision of survey statistics. The typical estimated between-interviewer variability for 10 personal interview surveys leads to an increase in the sample mean variance of 18% for an average workload of 10 interviews per interviewer. The same value of between-interviewer variability leads to an estimated increase in the sample mean variance of 48% for an average workload of 25 interviews (Groves 1989, p. 364).

A vast array of methodological and technological innovations in survey technology aims at minimizing between-interviewer variability. Data collection with computer-assisted telephone interviewing (CATI) controls flow and sequence of questions and facilitates supervision. In centralized interviewing facilities, furthermore, interviewers can communicate with each other so that interviewing techniques become more compatible over time. A number of studies have shown that centralized telephone interviewing generally can indeed reduce between-interviewer variability (Groves 1989, Tucker 1983, Groves & Magilavy 1986). On the other hand, this mode of data collection is typically accompanied by a high interviewer workload compared to personal surveys. Variance inflation of sample means in telephone surveys, therefore, might equal that of face-to-face interviews despite reduced between-interviewer variability.

Finally, collective strategies developed by interviewers might reduce variance but not necessarily yield 'better' data. Roth (1966 [1978]) observed coders who in response to a particularly ambiguous coding scheme agreed on a 'default' code. The 'shibboleth' of this group consequently became: 'Make it a 4!' Naturally, the supervisor who used between-coder variability as a device for monitoring reliability never knew what was going on.

[14] In contrast to most approaches to measurement of interviewer effects, the approach taken here does not depend on random assignment of interviewers to respondents. Although random assignment is highly desirable and should be part of all survey designs, it might not always be feasible or successful for a variety of technical or budget related problems (e.g. Tucker 1983).

Critics of the survey method point to the difficulty that standardized questions may be met with unstandardized answers. Fixed verbal utterances may change their meaning when meeting with variable linguistic horizons or communities of meaning (Brenner 1978, Deutscher 1978). The interviewer has the crucial role of communicating the meaning of a question under these variable conditions.

As Riesman (1958) puts it, the interviewers' task is to adapt the standardized questionnaire to the unstandardized respondent. Standardization of procedures may minimize interviewer variance but may increase bias due to poor comprehension or minimal memory search for relevant information by respondents (cf. Groves 1987). Interviewers who behave as "recording machines" (Hyman 1954, p. 68) might convey the impression that it doesn't matter what respondents answer as long as they do give an answer.

Prüfer & Rexroth (1985) experimented with interviewer training based on tape recordings of inadequate verbal behavior. They show that training can reduce variation in interviewer behavior considerably. They caution, however, that this reduction might not necessarily lead to 'better' data since interviewers simply reduce their verbal interaction with respondents to the indispensable minimum. However, Prüfer & Rexroth (1985) as well as Billiet & Loosveldt (1988) point to the crucial role of tape recordings as a control instrument. Differences in data quality between cases with and without tape recording might indicate that interviewers perform better when supervision is tight.

2.2. Perspectives on Interviewer Behavior and Experience

Some of the qualitative work which has been done leads researchers to conclude that "inadequate" interviewer behavior in terms of standardized procedures--such as changing the wording of questions and non-uniform probing techniques--yield more adequate data in terms of the substantive research interests (Pennef 1988).[15] As Brenner (1982) notes, interviewers engaging in these behaviors often get the best ratings by their supervisors (see also Pennef 1988).[16] Freeman and

[15]Some, but not all, of the field work presented by Riesman (1958) and Hyman (1954) confirms these observations, too.

[16]It is not clear on what supervisors based their judgement. I suspect that interviewers who are flexible enough to introduce 'creative' changes in the questionnaire based on their impression of respondents are also able to obtain high response rates. If Roth's (1966 [1978]) observations generalize to other settings, an entirely different mechanism might produce good ratings. Interviewers who are good at making up data that have a low variance in the sample mean, hence

Butler (1976) report that interviewers who obtained high ratings for interviewing skills from their team leaders produced actually more interviewer variance than those rated less skilled. One study found that more experienced interviewers used more unprogrammed speech and frequently altered the wording of questions (Bradburn 1983). Bradburn (1983, p. 314) hence concludes that flexibility of interviewers in adjusting their behavior and speech to the specific interview situation is a "real asset in carrying out surveys of the general population."

While mainstream methodological wisdom deals with sources of bias by eliminating the interviewer as an 'other,'[17] interaction theory predicts (1) that this does not yield unbiased data, and (2) it yields biases which we cannot easily identify as such. Argyris (1980) points out that the unilateral control exercised over the subjects of rigorous research may lead them to become either overly sensitive to researchers' behavior and try to please or second-guess them, or to develop uncooperative and "negativistic" attitudes towards the task and play games of minimal involvement (p.48-49). In contrast, Brenner (1978) argues that in the interviewing situation subjects are free to use their own relational standards and deviate from the task. Interviewers, however, are in a paradoxical structural position: they should "make the interview act without acting" (p. 138). Specifically, they have to remedy interactional trouble in the interview without getting involved with the substance of the reason for trouble since that might influence respondents' answers. When reference to formal, task, and etiquette standards fails, interviewers have "to give in to the idiosyncratic constructions of the respondent, that is, to not hear them" (Brenner 1978, p. 135). Hence, even when interviewers know that respondents misinterpret questions or give inadequate answers, the rules in a standardized interview may be seen as a normative prescription for subordination under respondents' task-inconsistent behavior.

On the other hand, survey practitioners are all too familiar with the frustration with interviewers who for no discernible reason skip questions or distort question content beyond recognition. In the final analysis, interviewing and interviewer training is a walk on a knife's edge between allowing 'beneficial' adaptations of the standardized questionnaire to the 'unstandardized' individuals and avoiding bias-generating distortions. While in this paper I cannot even begin to systematically evaluate social interaction in the interview, the nature of my data makes it possible to construct crude

---

those with realistic expectations about the sample distribution would then win the prize.

[17]Most radically in the self-administered CASI (computer-assisted self-administered interview).

measures which bear on the dynamics of interviewing.[18] Interviewer experience and its effect on data quality over time plays a central role in this analysis.

Rustemeyer (1977) finds that nearly one half of coded errors were "related to how well interviewers asked questions" (p. 342). Experienced interviewers make fewer errors than inexperienced interviewers but were more likely to change the scope of the questions. Consequently, she observes an increase of "invisible" errors with growing experience, i.e. errors that would not be detected in the coding stage.[19]

Groves & Fultz (1985) explain gender differences in initial refusal rates with male interviewers' higher job turnover. Singer et al. (1983) show that nonresponse rates decline initially with growing interviewer experience, but level off later on. They advance no explanation of this nonlinear influence of interviewer workload on response rate. Berk & Bernstein (1988) found no effect of interviewer experience on both nonresponse rates and data accuracy, comparing information on health expenditure obtained from interviews and medical providers.

The intuitive expectation about a trend in data quality would be that data quality increases with interviewers' familiarity with the mechanics of a difficult questionnaire and with concepts as well as the range of possible answers. After a while, one adjusts to the wide variety of incoming factual information and learns to handle classification problems. On the other hand, in a context of cross-pressures exerted by impatient respondents and an exceedingly difficult and long interview, interviewers might develop strategies to simplify their task; they might learn to skip tedious questions, discourage respondents to give exhaustive answers, and reduce probing. An application of an organizational behavior perspective to the survey setting leads to three hypotheses:

(1) the more 'rigorous' research is organized, i.e. the more interviewer involvement in the substantive goals is avoided, and bureaucratic control over the data collection process is

---

[18]Ideally, research should combine the approach taken here with the interaction coding technique developed by Marquis and his associates (Marquis 1969). Verbal behavior by respondents and interviewers could then be linked to outcomes in terms of data accuracy in a systematic way. In the section on data editing (see below), I provide some anecdotical evidence.

[19]These results are based on mock interviews. It is not clear how they relate to real life survey conditions. Inexperienced interviewers might feel considerable pressure under these circumstances, while experienced interviewers might not take the task as seriously as a true interview. These are pure speculations, however, and the opposite argument could well be made.

developed, the more interviewers tend to maximize their own utility at the expense of research interests (sloppy work, outright cheating, reduced output).

(2) These processes (alienation) in a survey organization lead to an increase in 'invisible' errors: interviewers learn to avoid visible traces of 'inadequate' behavior.

(3) Supervisors will not be aware of these processes because they concentrate on visible sources of error. They will reinforce interviewers' 'strategic' conformity to the rules while not being able to reward 'authentic' research goal-oriented behavior (Roth 1966 [1978], Argyris 1968, 1980).

While I will not be able to test these hypotheses explicitly, I will address some related questions. The approach described above suggests a tension between appearances and meaningfulness of data.[20] Are these two aspects of data quality really different in terms of response effects influencing them, if any? Do invisible errors increase over time due to alienation processes while visible errors decline due to unintended learning processes? Survey researchers should be concerned with these questions both for evaluating and monitoring survey data quality.

[20]My study of the survey methodology literature did not provide me with an adequate terminology tapping these two aspects of data quality. By analogy, they might be viewed as syntax and semantics. A sentence with a perfect syntax might be utterly meaningless when the semantics are wrong and vice versa. Although both are necessary for assessing the quality of a text, they are independent concepts.

III. DATA AND MEASUREMENT

3.1. The Life History Study

As part of the German Life History study, the Max Planck Institute for Human Development and Education (Berlin; principal investigator: Karl Ulrich Mayer) in cooperation with Infratest (Munich) collected life course data of 2004[21] men and women born between 1954-56 and 1959-61 from November 1988 until October 1989 (Brückner & Mayer, 1995). A computer-assisted telephone interview (CATI) was designed to facilitate and support retrospective data collection. The questionnaire included questions on migration history, education, job career, family of origin and own family, health, as well as voting behavior and standard demographic variables (see Figure 1 for an overview over most of the content and the structure of the questionnaire).

The CATI technology offers opportunities to incorporate simple consistency checks, personalized questions, highlighting of important stimuli, and monitoring the ongoing interview process.[22] Considerable effort went into programming, interviewer training, and supervision. A two-day training session with mock interviews provided interviewers with a basic familiarity with the questionnaire and measurement concepts. In addition, each interviewer obtained printed material pointing out critical parts of the schedule, especially ramifications of the German educational system with related concepts and classification problems in the migration history.

Respondents were called during the evening (between 6 and 9 pm). On the average, 8 interviewers worked together in one shift. For the first 7 months of the survey, two supervisors were responsible for scheduling and monitoring of interviewers. In the remaining 5 months, one supervisor was felt to suffice.

All completed interviews were reviewed in a two-stage editing process. Extensive consistency checks between and among the available life domains were made, using tape recordings (available for about two thirds of all cases), as well as call-backs to respondents for clarifying remaining inconsisten-

---

[21]The actual number of completed interviews was 2008. Four interviews were conducted by two senior researchers and two supervisors and are therefore excluded from the analysis which is based accordingly on 2004 cases and 24 interviewers.

[22]For example, answers given earlier in the interview would be used to ask questions which incorporated the specific context such as "And when did you start working as a self-employed carpenter?"

cies.[23] Having worked in the previous life course survey (cohorts 1919-21) as a research assistant, I joined the editing group when the study was about halfway through the field and served as editing supervisor and liaison between the institutes in Berlin and Munich for the remaining field time. Occasionally, the editing group in Berlin would summarize feed-back for interviewers and communicate them to the senior researchers in Berlin and Munich. In July 1989, for example, we compiled a series of tape recordings with inadequate or incomplete career histories and used them for an additional training session with the interviewers, trying to clarify the concept of a job change.

Tape recording of interviews for editing purposes in surveys is a rare luxury; it is expensive, and, furthermore, not all respondents agree to having their interview recorded. However, the tapes provide an invaluable data source for editing and field control. Despite the advantages of CATI technology and the tapes, call-backs were necessary in about one quarter of all cases. The editing process itself is very expensive and takes a long time.[24] An analysis of sources of response error might indicate possible improvements in the data collection stage and possible ways to reduce costs in the editing stage.

Interviewers with superior performance were selected from the pool of interviewers working in the telephone studio of the commercial department of the cooperation institute. All had previous experience with CATI systems; seven had worked in the previous life course study of the 1919-20 cohort which used essentially the same questionnaire. Interviewers were paid by the hour and not, as customary in face-to-face surveys, per completed interview. The overall average workload was 84 cases per interviewer. However, five of the total of 24 interviewers gave up or were reallocated to the commercial branch after one or two months because of personal reasons or a below-standard performance in 10-28 completed interviews. These five interviewers completed 83 cases (average workload 16.6). The remaining 19 interviewers completed between 41 and 184 with an average interviewer workload of 101. Compared to previous parts of the Life History Study, where face-to-face interviews were conducted with an average number of 6-7 cases per interviewer, this workload is extremely high. The high workload in the study under consideration here reflects organizational differences between the two modes of interviewing in the administration of a national survey.

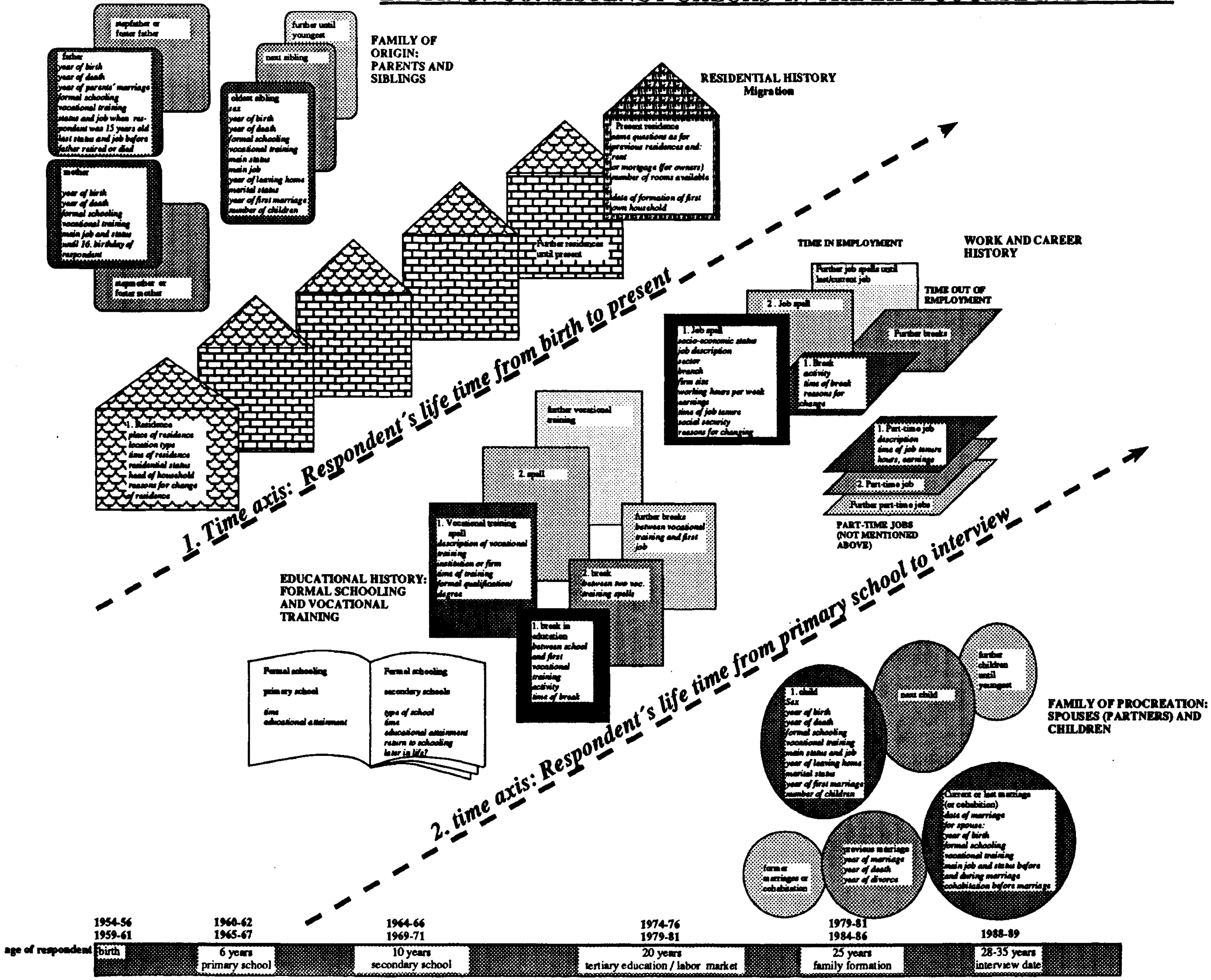
---

[23]For a more detailed description of the editing process, see below and Nuthmann & Brückner (1995).

[24]Mayer & Huinink (1990) report that for the first part of the Life History Study (birth cohorts 1929-51), editing cost about the same as the interviewing itself. For the data analyzed here, editing (including data entry of corrections) took two years with a staff of one full-time and five to six part-time employees.

Figure 1: Overview of Questionnaire

EDITING: CONSISTENCY CHECKS IN THE LIFE COURSE INTERVIEW



Source: Max Planck Institute for Human Development and Education, German Life History Study, Cohort 1954-1956, 1959-61

Centralized data collection with a limited number of work stations and high training costs associated with the CATI method and the difficult instrument lead to the extreme workload differences reported above. A low refusal rate[25] and overall good cooperation of respondents should increase interviewers' confidence in the study and put them at ease with their task so that positive effects of interviewer experience may be expected (see, for example, Converse & Schuman 1974). However, interviewer bias may be exacerbated and hence offset a positive effect of interviewer experience on response rates and data quality.

The analysis of effects of increasing workload presented here may inform future research facing the trade-off between costs and benefits of interviewer experience. Groves & Magilavy (1986) report average interviewer workloads of 12-58 in 9 telephone surveys in the U.S., while in 11 telephone opinion polls Tucker (1983) analyzed interviewers had workloads between 22-38. While not as high as in the study under consideration here, these figures do suggest a trend towards higher workloads than those customarily aimed for in face-to-face survey research. For example, Brückner (1989) cites 7 cases per interviewer as a desirable figure.

No personal data on interviewers are available because of concerns with confidentiality. Since in terms of age and education there was little variation among the interviewers anyway, this is not a big loss. Rather, the analysis focuses on job experience and trends in performance as well as on the effects of work place organization. Given the small number of persons, I introduce interviewers as individuals rather than as categories in the models where appropriate.

Before I proceed to a detailed description of variables used in my analysis, I describe techniques and some of the problems emerging in the editing process. Part of my analysis rests on the assumption that the edited version of the dataset comes nearer to a 'true' value of the number of jobs held over the life course than the unedited version. The following attempts to justify this assumption.

3.2. Data Editing

Although data editing belongs to everyday life in social research, the documentation is usually sparse and leads a modest life in footnotes of published articles, if any. It is always heavy with sometimes 'subjective' looking decisions on how to treat problematic features in the data. One does create rules, of course, but many problems surface in the editing precisely because they are idiosyncratic and unanticipated in the design. Unanticipated and idiosyncratic problems trigger the

[25] 9.1% in the study used here.

creation of ad hoc rules. As a result, the procedure may have an 'unscientific' flavor. What I hope to show in this analysis, among other things, is that data editing may reduce systematic error which we might otherwise ignore.

What is the justification for editing? What are the standards against which data quality is measured here? As always, there is an easy and a difficult answer. The easy answer is that we evaluated the obtained data on the basis of the concepts we set out to measure. Max Weber (1949) insisted that the social scientist must aim for "analytically ordering empirical reality in a manner which lays claim to validity as empirical truth" such that the correctness of a scientific proof could be acknowledged "even by a Chinese" (p. 58, emphasis in the original). Viewing measurement concepts such as a job spell as ordering empirical reality in way that makes sense for scientific analysis helps clarify why edited data can come nearer to a 'true value' than what was obtained in the interview, even though editing is not one iota less the outcome of a social process than data collection itself.

A concept which might make sense to a Chinese social scientist does not necessarily correspond to the way respondents order empirical reality when thinking or talking about their lives and experiences. The usefulness of survey data critically depends, therefore, on communicating the particular ordering of empirical reality undertaken by the researcher to all respondents even or especially when the latter may find it tedious, threatening, unfamiliar, or strange.[26] Editing, then, is neither more nor less than an attempt to compensate for shortcomings of the communication process in the interview by reordering and reconsidering data. The more difficult question is: how can it be done? Below, I try to trace the process using a few examples.[27]

One of the most important aspects of editing is not to confuse consistency and conformity. People experience all kinds of 'unusual' life trajectories. Rather, the editor looks for logical impossibilities, e.g. negative durations. In regard to job trajectories, the operationalization of the concept of a 'job spell' is key to the editing process. A change in occupation, position, hours worked, and (local unit of) the firm are all markers for the beginning of a new job spell. Hence, a person who starts out as a clerical employee in public administration but lists as her current form of contract 'civil

---

[26]Indeed, respondents often expressed astonishment at the detailed questions and wondered what one could possibly do with all this information. Some requested a printout of the data or a copy of the tape recording since "by themselves," they "would never be able to put things together" like in the interview.

[27]The interested reader can find more information on the specifics and rules in Tölke (1989), Brückner E. (1990, 1994), Brückner E. (1993, Vol. III). The handbook for data editing for the present study is published (Nuthmann & Brückner, 1995).

servant' should have at least two job spells. The transition from employee to civil servant corresponds to the end of one spell and the beginning of the next.

Similarly, a person who moves from one end of the country to the other should experience a corresponding job change, if only in terms of the establishment of the firm. Of course, people can commute over extended periods of time and long distances. In most cases of inconsistencies between the migration and career history, the tape recording of the interview helped clarify the problem. Otherwise, the respondent was contacted. If the respondent could not be contacted, the data were left in its original form and a note was made to mark these cases as potentially inconsistent.

The interviewer has an important role in keeping the respondent alert to the different concepts and dimensions involved in the job history. At the end of each job spell, the first filter question inquired about the next activity (see Figure 2 for an overview of the filtering in the career history). If the respondent continued to work, he or she was first asked whether the next job was in the same firm (local establishment). If yes, she or he was asked whether the firm had changed its size or the industry it was operating in.[28] Next, she or he was asked if her occupational status had changed. These filter questions are critical for a complete and correct coding of the job history.

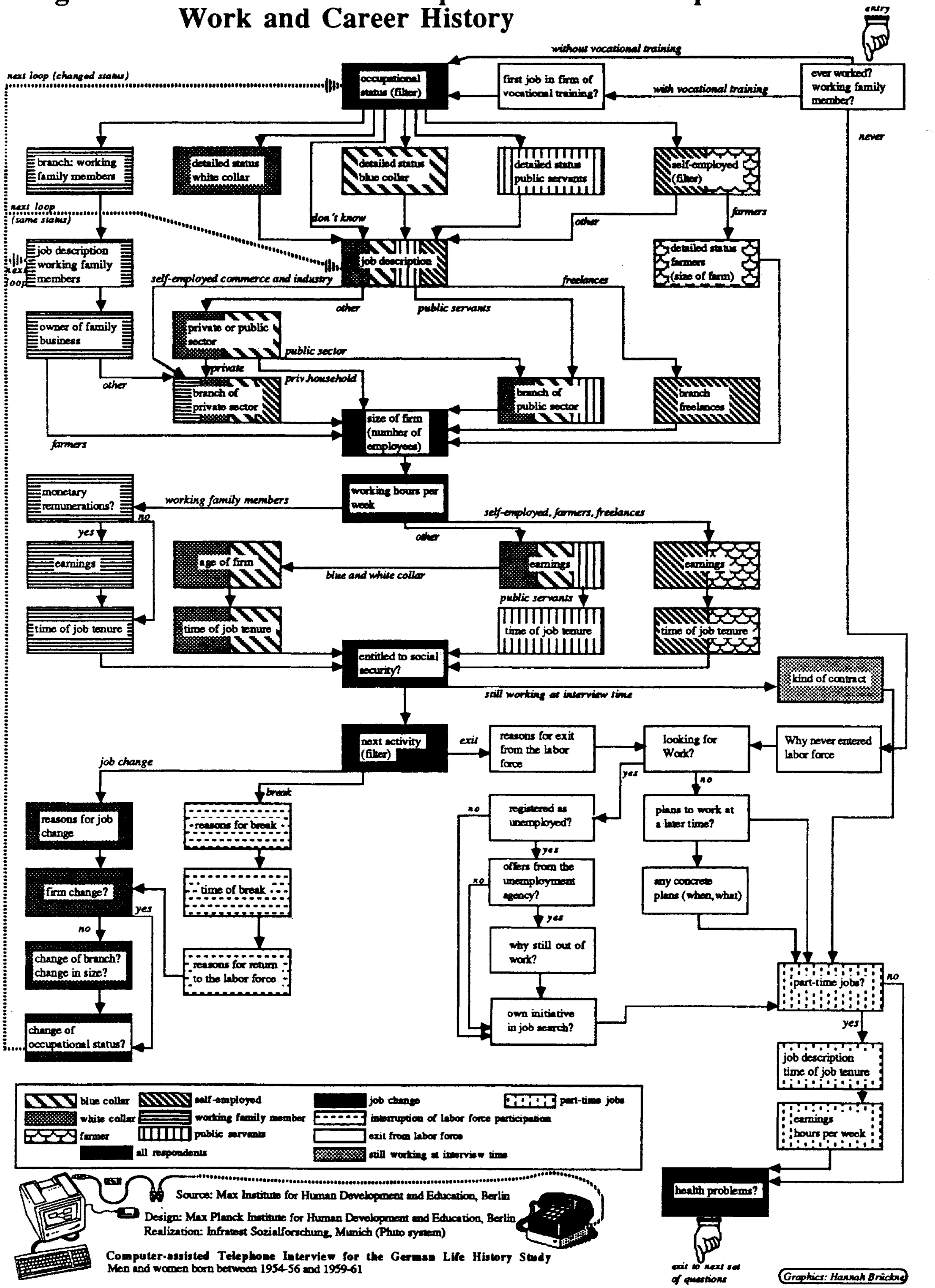
In one case the tape recording tells a story about what can go wrong: the respondent (R) had worked as a retail salesman in his first job. After a while, he obtained an internal promotion. When asked by the interviewer (I) whether the firm had remained in the same industry, he answered:

R: *"Yes, that was always the same."*
I: *"You always worked in the same firm?"*
R: *"Let's say...uh, yes"*
I: *"And you have been working there until today?"*
R: *"I am still in the same business"*

The result of this interaction was that the next and last job spell contained the rest of this man's career, and listed the same data as the first job. Only the occupation was different. We found out later that a whole series of career moves had been omitted. Originally, we had gone back to the tape because both jobs listed the same occupational status while the first job ended with a promotion.

[28]If size and industry had remained constant, the corresponding questions in the next sequence were skipped. The information already given in the preceding job spell was automatically duplicated and put in the appropriate place. This procedure saved valuable interview time. Also, an unnerving repetition of questions and information already obtained was avoided. Computer-assisted interviewing makes complicated skipping pattern such as this possible.

Figure 2: Filters in the Computer-assisted Telephone Interview: Work and Career History



In addition, the respondent had taken courses in business leadership and personnel management. Clearly, the interviewer discouraged the initially very cooperative respondent to go into further detail about his career.

This is an extreme case, reported here for the sake of illustration. Generally, 'output reduction,' if any, by both respondents and interviewers was much more subtle. More often, respondents misunderstood questions or got confused. Interviewers did not always understand that. Sometimes they would choose to ignore respondents' misperceptions or confusion to avoid embarrassment or additional delay.[29] Hence, inadequate or inconsistent answers were coded 'at face value' without further inquiry. The task of the editing group was to detect and mend these subtle 'breaks' or gaps in the data. A lot of intuition is involved in this process, and additional information from the whole interview has to be used for the evaluation of invisible errors.

This description has concentrated on the correction of 'invisible' errors because I asserted earlier that editing produces data which come nearer to a 'true' value than the raw data. I would like to add, however, that the correction of 'visible' errors is not a separate process--one cannot correct one without evaluating the other. We find missing data and chronological inconsistencies in dates across the life course for a variety of reasons; correction of these errors presupposes a careful reconstruction of both the interview process and features of respondents' life courses. In sum, interviews for the most part contained most of the desired information--but not always in a workable format. Expensive and tedious procedures had to be developed to get the data ready for analysis. This is disappointing, given that simple consistency checks and various visual helps were already built into the CATI design. In the analysis presented below, I attempt to trace the reasons for this.

3.3. Design

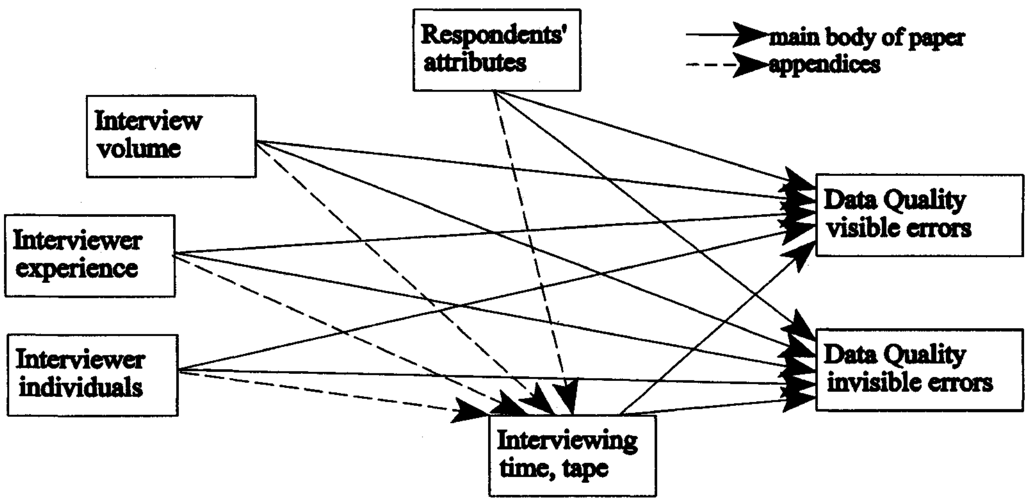
The design of the empirical analysis can be described mainly in terms of three strategies. First, I develop two indicators aiming at different aspects of data quality. Second, by introducing individual interviewers as explanatory factors, I address the question of potential interviewer effects on data quality in the light of the high work loads interviewers had in this study. In addition, respondents' attributes serve as proxy variables for processes of information retrieval processes, and a measure of life course complexity reflects potential difficulties with the questionnaire. Third, by comparing effects

---

[29]Often, however, interviewers made a note for the editors to indicate potentially problematic parts of the interview.

of independent variables across models, I evaluate the usefulness of a more differentiated conceptualization of data quality.

Figure 3: Causal Model



Conceptually, when I talk about 'interviewer effects' in the following empirical analysis, I mean differences between individual interviewers. This emphasis on individuals is possible and important in the present study precisely because interviewers did so many interviews. That is, the potential of an interviewer bias increases as work loads become higher. Figure 3 sketches the causal model.

3.4. Variable description and specific hypotheses

3.4.1. Data Quality as a Multidimensional Concept

Visible errors. The elimination of visible errors is crucial for the successful application of dynamic analysis techniques such as event history analysis since one missing datum in an event sequence may necessitate the elimination of the entire case from the analysis. In addition to and in support of the manual editing, a computer-assisted editing system was designed to detect remaining errors and inconsistencies such as gaps and overlaps between events in the chronological sequence (Brückner H., 1995). For the measurement of visible errors, the program was used to scrutinize the unedited data, and selected error types were listed casewise and matched to the respondents.

Missing data on event start and end dates, and gaps and overlaps in the event sequence of migration history, education and career history, in the data as recorded by the interviewers are summed as a measure of overall data quality. Each missing was counted as a visible error (that is, when both month and year were missing, this counted as two errors, so that one event can have a maximum of four visible errors due to missing data). Moreover, when two adjacent events had only missing data on the dates, this was counted as an overlap by the check program, so that a certain inflation of visible errors results. As I said above, the program was designed to detect errors which were overlooked during editing and data entry rather than for counting errors. On the other hand, two adjacent events with missing data are really a more serious data problem than one surrounded by others who, at least, had some dates on them.

Table I reports the frequency of visible errors in education and career history by type and case and by type and events. More than third of all respondents did not know at least one month, and every fifth respondent did not know at least one year. Overall, 40% of cases had at least one missing on at least one month or year. A break down of missing dates by spells looks much less dramatic: about eight percent of events had a missing month, and about three percent had a missing year on either event start or end.

Of course, years are easier to recall than months. Considering that respondents were asked to recall dates over the whole period between entry into elementary school to interview date, amounting to 22 up to 28 years, depending on age, that might not seem too bad. Still, from the editing process we know that interviewers often failed to probe for more information. Respondents sometimes needed more time to reconstruct the dates than interviewers would give them; and finally, the easiest way to get through especially the early event history was, of course, to modify the questions and ask respondents only for the year.

In these instances month or year would appear as missing but it is not at all clear that respondents could not have come up with a date, given more time or more probing or simply if they had been asked for it. A great deal of work went into correcting all these "visible errors". The check program described above was designed to assist manual editing as it turned out that in a considerable number of cases such errors had been overlooked by the editors (about 10% of all cases showed visible errors after the second round of editing).[30]

Table I: Visible Errors in the Life History Sequence.

Missing Data, by case and spell, broken down by spell start and end, in percent:

	Percent Missing: by case	by spell
Start month	37.7	8.1
Start year	20.0	2.9
End month	35.6	7.8
End year	19.3	2.8

Overlaps and gaps, each spell compared with all other spells in one case, casewise, in percent:

error count per case	overlaps	gaps
0	7.9	15.6
1	16.5	31.9
2	14.7	29.5
3	14.0	15.6
4	12.0	5.1
5	8.9	1.8
6	6.7	.4
>6	19.1	.2
N	1987	16,192

Numbers based on 1987 cases, German Life History Study, Birth Cohorts 1954-56 and 1959-61.

Similarly, the great number of overlaps--only 8% of all cases does not have at least one overlap--is to explain partially by interviewers' tendency to skip the questions about starting dates and put the ending date of the last event coded. This might not seem too severe a transgression. It could create, however, great confusion in the event history where the chronology of life events was different from the sequence of questions because of the questionnaire structure. Hence, when respondents' life course 'oscillated' between school, work, and other activities, asking and coding all dates was crucial to avoid confusion and overlaps.

---

[30]With the help of the program, all of these were eventually eliminated. The rules and procedures are further described in Nuthmann & Brückner (1995).

Summing up over all these errors leads to a mean of visible errors of about 12 per case, reported in Table II. For many purposes, data on only some of the events listed in Table I are required. However, one of the main thrusts of the survey was to acquire an as complete as possible documentation of the life course; a study interested in the relationship between migration and career history, for example, would end up with less than half of all cases when deleting only those with missing data on the event dates. This might still be a sample of acceptable size when not too many cases have also missing data on the co-variates, and if it resulted in a random selection.

If, however, as I will show below, the amount of visible and invisible errors is related to how complicated respondents' life courses are, deletion of missing or inconsistent cases might soon result in a selectivity bias on the dependent variable. Assume, for example, that we obtain lower data quality for more mobile people. An analysis of mobility with deletion of missing data would then select against mobile respondents.[31] In addition, inflation of calculated durations due to overlap among events is likely to increase with increasing number of events.

Invisible errors. For the analysis of 'invisible' data quality, I use the number of job spells before and after editing as well as the difference between the two, assuming that the editing process entails an approximation to a 'true' value of this variable. The focus on jobs was motivated by the central position labor force participation and career trajectory have in the 'real world' as well as in research. A systematic increase in added job spells over the field time may indicate unintended learning processes leading to output reduction. Each additional job spell triggers a long set of questions (dates, firm size, industry, sector, salary, type of position, occupation, social security coverage, and hours worked). Gibson et al. (1978) found that the number of acute health conditions reported dropped an alarming 20% when a special supplement of additional questions about such conditions was added to the survey (cf. Martin 1983). Table II shows that in the modal case, no job spells were added; however, up to ten spells were added for some careers.

Multidimensionality. Invisible errors are conceptualized in this paper as a way to evaluate the meaningfulness of data with respect to the goals and concepts of a survey. This leads naturally to the

---

[31]Anderson et al. (1983) propose estimation procedures for missing data in surveys which are more efficient than simple substitution of the mean of a given variable. Note, however, that these techniques assume no association between patterns of missing data and the dependent variable to be analyzed with the estimated data. Possibly, techniques are available to overcome the latter problem. Data edition as described above seems preferable since we incorporated information from the whole interview and tape recordings when imputing data. I doubt that the complex considerations involved could be matched by an estimation process.

Table II: Means, Standard Deviations, and Description of Variables. German Life History Study, Birth Cohorts 1954-56 and 1959-61 (N=1987).

Variables	Range	Mean	Std.
Dependent variables:			
Visible errors (Number of errors in the event history sequence found by the check program)	0-56	12.33	7.36
Invisible errors (Number of Jobs added)	0-10	.23	.66
Interview characteristics and Volume:			
Log of Time (in minutes)	3.18-5.30 (24-200)	4.15 (66.90)	.32 (22.81)
Log of total number of records (raw numbers)	3.18-5.12 (24-168)	4.02 (58.23)	.27 (16.52)
Log of number of events checked (raw numbers)	1.39-3.74 (4-42)	2.57 (13.73)	.32 (4.59)
Number of job spells before editing	0-13	2.78	1.73
Number of job spells after editing	0-14	2.89	1.77
Respondents' characteristics:			
Age (in years)	26.92-35.75	31.33	2.70
Gender (1=female)	0-1	.49	.50
Proportion living with partner	0-1	.70	.46
Number of children under 6 in household	0-3	.51	.72
Schooling: highest degree obtained	2-6	4.13	1.24
Life course pattern complexity	0-9	3.26	1.47
Life course pattern relative frequency	1-7	4.71	2.24
Log of time elapsed since last life course transition (in months)	1.10-5.72 (0-305)	3.99 (72.01)	.83 (51.36)
Interviewing:			
Interview completed in two sessions (1=two sessions)	0-1	.23	.42
Interview completed in three or more sessions (1=three or more sessions)	0-1	.02	.15
Number of interviewers working in shift	1-16	8.45	2.72
Weekly productivity	1-9	3.25	1.81
Tape recording (1=yes)	0-1	.75	.43
Interviewer experience (number of interviews completed before the ith interview)	0-183	58.73	44.75

question whether this second aspect of data quality is really different from what I called here 'visible' data quality. The zero-order correlation is positive but low with r=.0535 but a bivariate statistic may be misleading; in any case, it does not prove that they are not generated by the same processes. Are the two types of error just two indicators for the same underlying concept (i.e. the propensity to produce accurate and complete information)?

In this case, we would expect effects on data quality to operate the same way for both types of errors. Do interviewers who produce on the average many visible errors also produce many invisible errors? In section 4.1., I describe differences between interviewers with respect to their mean on visible and invisible errors. Finally, in section 4.3.3. I will compare effects of independent variables across models. I will show differences and similarities in the effect of interviewers, respondents, and the interview situation on both types of error.

Common sense would predict a general decline of errors over field time, since interviewers become more acquainted with the questionnaire and more confident in dealing with difficult respondents. Critics of the survey method might predict a decline in visible errors but an increase in invisible errors as interviewers learn to reduce output but hide obvious traces of inadequate behavior. A tendency to change question wording with growing experience might contribute to an increase in both visible and invisible errors. Table II reports descriptive statistics of dependent and independent variables for both error count models.[32] The following section describes independent variables and hypotheses about their effect on data quality.

3.4.2. Independent Variables

The variables described below aim at tapping main concepts identified in previous research as determinants of data quality: (1) memory retrieval processes, both in terms of salience and remoteness in time; (2) respondents' and interviewers' ability to deal with the questionnaire; (3) potential distractions from the interaction process, both with respect to interviewers' work place and respondents' home; (4) length and volume of the interview as a measure of efficiency or 'flow' of interviewing; and (5) interviewer experience and productivity as indicators of organizational dynamics.

---

[32]Table II and all the following results are based on 1987 cases. 17 cases were deleted because of missing data on the co-variates. I found no indication that these cases differ from the included cases in terms of the variables used here.

Interviewer and interviewing:

Individual interviewers: For each interviewer a dummy variable indicates if a particular interviewer had conducted the interview under observation. As a baseline for the comparison of individual interviewers I use those 5 interviewers who left the project early and completed less than 30 interviews (N=84). This procedure is somewhat unusual and it might seem preferable to chose the interviewer with the highest number of completed cases as a baseline for the analysis. However, it enables to me to show differences between interviewers with exceptionally high workloads to those who left the study for whatever reasons.[33]

Interviewer experience: A count variable indicates the number of interviews conducted by a particular interviewer before the current observation. That is, experience here stands for tenure in the survey under consideration since data on previous employment as interviewer are not available. Tests for interactions and curvilinearity aim at clarifying the mechanism by which experience operates.[34]

Tape recording: Provided respondents agreed, interviews were recorded. These tapes, available for two thirds of all interviews, are an important resource in the editing process but also provide a wealth of qualitative data. For the present study, they are used as measure of control over interviewer performance. Kahn & Cannel (1982) as well as Prüfer & Rexroth (1985) underline the crucial role of recording interviews. Interviewers who know that an interview was recorded may adhere more closely to the questionnaire, and avoid negatively sanctioned behavior such as leading or directive probing, and skipping questions. To the extent that this kind of behavior has a negative effect on data quality, recorded interviews are expected to be less error-prone than those which were not recorded. On the

[33]Of course, this forces the impact, if any, of these 5 interviewers to be equal in the model. Table VI shows the results of the models if the interviewer with the highest workload is taken as the default category. See below for the comparison of the two models.

[34]We know which interviewers worked in the previous part of the study (birth cohorts 1919-21). Since I introduce dummy variables for individuals, I cannot simultaneously flag those with prior tenure with the study because a dummy variable would be fully determined by individuals' dummies. When omitting the latter, prior experience has a negative effect on both error counts which is most likely due to a selection process in that (1) the best interviewers are likely to be selected for a continuation and (2) motivated and enthusiastic interviewers are more likely to accept further employment in the difficult study. However, when using the results to identify the 'best' interviewers, not all of them had worked in the previous study.

other hand, editors might be more likely to add job spells in recorded interviews because more information is available. In the analysis of invisible errors, therefore, the two effects may cancel each other out.[35]

Weekly productivity: Number of interviews completed in a period of seven days before the date of the interview under consideration, ranging between one and nine with a mean of about three (the minimum is one because the current interview is included). When controlling for experience, weekly output indicates routinization effects associated with times of high productivity. Such effects may be both positively and negatively associated with data quality. A positive effect could be interpreted as an increase in efficiency due to an "automation" in dealing with the difficult instrument. A negative effect, conversely, could be interpreted as an effect of boredom and being exasperated with the job.

Interruptions: Interviewing was limited to the evening hours. Sometimes interviewers or respondents had to interrupt because of various time constraints. 464 interviews were conducted in two sessions, 45 in three or more sessions. Respondents may have more time to reflect on their past during the interruption, or to retrieve documents and resumes. On the other hand, a disruption in the retrieval process might increase recall error. Interruptions might also put additional strain on interviewers' task. Furthermore, respondents might be less cooperative after having had time to reflect on the situation. Therefore, interviewers might accelerate the pace in the second session in order to avoid additional interruptions.[36]

Number of interviewers working together in a shift: In order to control for disturbing influences on the interviewer side, the number of interviewers working on the day the interview was completed is introduced. As Table II shows, on the average about eight interviewers worked together in a shift. Everybody who visits a telephone studio during working hours can observe that it is a very lively environment, to put it positively. People constantly talk, dial, try to attract supervisors' attention, or walk around in the search of coffee or new phone numbers. The machines and telephones were placed in

---

[35]In cases without tape recording, significantly more call-backs to respondents were necessary. However, not all respondents could be contacted, and some refused to answer further questions.

[36]In addition, presence of small children in the household is introduced to control for possible distractions and disturbances during the interview. For this and the potential presence of other persons, see below, under living arrangements.

carrels which provided some screening, but basically everybody worked in the same room. Accordingly, a greater shift size may have a negative effect on interviewers' concentration and hence data quality.

Interview characteristics and volume:

Number of events (spells) checked by the editing program: The life course interview comprised 4-42 spells with a mean of about 14. Obviously, the more events that are coded in the interview, the more errors can be found by the check program. In the analysis of visible errors, one specific hypothesis worth testing is whether the number of events is roughly proportional to the number of errors found or whether an increase in events leads to a disproportional expected increase in the error count due to 'sloppiness' of interviewers and/or 'output reduction' by respondents. For the analysis of invisible errors, I have no a priori hypothesis in regard to this variable but retain it in the model for the convenience of comparison.

Number of job spells (before editing): For the analysis of invisible errors, number of job spells serves as an indicator of 'output reducing' tendencies among interviewers.[37] In the raw data, interviews had between 0 and 13 job spells with a mean of 2.78. After editing, the mean was 2.89, amounting to an about 4% increase in the average number of jobs during data editing. With each additional job came a large set of additional questions. The more jobs were coded, the more interviewers might discourage respondents to report even more. Of course, respondents themselves might lose their patience with the tedious task and try to shorten the interview. A squared term is used to test for a curvilinear relationship. In the model for visible errors, the meaning of this variable is different since omitted job spells would rarely show up as a true gap in the event sequence; rather, an effect of additional jobs on visible errors net of number of events may tap the concept of salience and hence improved recall. The effect is expected to be negative, confirming a hypothesized greater salience of events associated with labor force participation as compared to events in the migration history or education.

---

[37]This variable can be seen as a characteristic of respondents. In the context of this analysis, however, I conceptualized it more in terms of being the outcome of the interaction process in the interview.

Interview volume (total number of records stored in data base): This measure includes all events as well as persons (family members for which specific questions were asked) and number of lines in text records with answers to open questions.[38]

Interview time: One of the nice features of the CATI technology is that interview time can be measured very precisely by using the computer as a 'stop watch.'[39] Time is a proxy variable for two distinct processes in this analysis. First, respondents who are insecure and have problems answering questions and remembering dates might take longer to complete the interview. Hence, a positive correlation of time and number of errors is expected. Secondly, time also might indicate interviewer behavior such as skipping questions and pushing as well as respondents' tendency to answer quickly, without reflecting. Here, a negative correlation is expected. The direct effect of time may therefore be a net effect of these two relationships. Interview time may be conditional on the amount of information recorded in the interview: when holding constant volume, additional time may enhance data quality since respondents take more time to recall information. Also, additional time while holding constant volume might indicate that interviewers ask all questions and take time to explain concepts or do not rush respondents.

Respondents' characteristics:

Gender: Within the context of the present study and its focus on factual data, respondents' gender might serve as a proxy for the unmeasured intervening variable 'salience of events.' The importance of salience has been shown by previous research (Sudman & Bradburn 1974, Cannel et al. 1981). Arguing from a socialization perspective, career histories might be less salient for female respondents who identify instead with their family and associated events. On the other hand, experiences with the previous life course studies conducted by the research group showed in that in some cases women

---

[38]The questionnaire included a considerable amount of semi-open questions, i.e. questions for which a number of answer categories (including 'don't knows' and 'refused' were provided but also a space for 'open' answers when fit between categories and answers was not obvious.

[39]Interview time is especially important with respect to cost efficiency. Here, I concentrate on aspects of data quality rather than on cost efficiency although some aspects of the present study may have implications for the question of how to get the best data per time unit.

were the 'keepers' of biographical memory within their families. Male respondents frequently consulted their wives during the interviews when they were asked about their own lives (Dex 1991).[40]

Age: Research has shown that recall problems increase with the span of time to be remembered (e.g. Wagner 1989). Hence, it might be more difficult for older respondents to remember events more remote in time than for younger respondents. In the data used here, the maximal age difference between respondents is only seven years; these seven years between 28 and 35, however, comprise a time in peoples' lives when a lot of things such as family formation and 'settling down' are likely to happen. These cohorts have been termed the 'postponed generation'; after participating in the expansion of the educational system, they experienced extended search for a job or a 'vocation' under unfavorable labor market conditions. For the older cohorts, this search phase may already be past and 'forgotten' while the younger cohorts are more likely to be still concerned about it. Thus, age should increase the error count.

Living situation: A dummy variable indicates whether the respondent is currently living with a 'significant other' (regardless of whether they are married or not). About 70% of all respondents are living with a partner. The (potential) presence of 'third' persons during the interview was of concern in previous research (e.g. Hoag & Allerbeck 1984). Third persons may interrupt the communication, distract respondents, or inhibit respondents' frankness in answering questions. In addition, the number of small children (under the age of 6) in the household may tap potential distractions. Interviewers complained occasionally about respondents being distracted by yelling or crying children in the background.

Schooling: The highest degree obtained by respondents is measured on an ordinal scale ranging from 2 to 6 with the following categories: 2=less than Hauptschule (10 years or less, without certificate), 3=Hauptschulabschluss (8-10 years, with certificate), 4=Mittelschule (middle school degree, 10 years), 5=Fachhochschulreife (12 years, qualification for professional colleges), 6=Abitur (13 years,

---

[40]Respondents' gender plays a role in the interaction theory of interviewing. Interactions between interviewers' and respondents' gender may affect responses, although the evidence provided by previous research is ambiguous. About half of our interviewers were male. A test of different combinations of interviewer-respondent in terms of gender coded as dummy variables (male-male, female- male, and male-female compared to female-female, as well same-sex dyads compared to others) for both error types yielded no significant coefficient.

qualification for universities). There are two reasons to use the degree categories rather than duration; if schooling is seen as indicating cognitive ability, degree gives a better measure within the context of the German educational system, which is both highly stratified and standardized (Allmendinger 1989; for a description of the German system, see also Kalleberg & Witte, 1992). Different degrees are obtained at different types of schools which emphasize different kinds of training.[41] Degree obtained may also be seen as facilitating the access to 'orderly' careers.

Complexity of life course patterns: Transitions between formal schooling, vocational training, and labor force participation were counted.[42] These three 'domains' were separate segments of the questionnaire (see Figure 1). Respondents were first asked about all spells in formal schooling, followed by the sequence of all vocational training spells, and then finally about their job history. Especially in the cohorts under observation, considerable oscillation between these domains or 'life course disorder' may occur.[43] Difficulties reporting dates related to these events may result from partitioning chronological sequences into different strands. For the purpose of my analysis, the number of transitions between life domains before editing is counted. As Table II shows, 0-9 of such transitions occurred with a mean of three. A score of zero means that the respondent never worked or received any kind of training after leaving formal schooling.

Relative frequency of life course patterns: Interviewer expectations have been a concern of researchers since Hyman's analysis of interviewer effects (Hyman et al. 1954). He showed that interviewers'

---

[41]I test whether duration has an additional effect on data quality when controlling for degree as well as for linearity of the degree effect.

[42]For this purpose, adjacent spells in each of the three domains were collapsed so that the sequence school1-school2-work1-work2-training1-training2-work3 became school-work-training-work. Note that this operationalization is strictly modelled after the data collection process to measure difficulties associated with the fragmentation of chronological sequences.

[43]See also Rindfuss et al. (1987). My variable is different in that I coded simply the number of transitions regardless of the duration of the constituting events. Furthermore, the period under observation is longer in my analysis (from first entry in the educational system in the period between 1960-65 until 1989). Transitions between labor market and activities such as homemaking, unemployment, and other 'gaps' after the first entry into the labor market are not considered since they posed no special problems in the data collection (i.e. the questionnaire proceeded simply chronologically). Also, the great heterogeneity noted by Rindfuss et al. produces not only computational problems but also a large number of unique or rare patterns which are not really informative for the present purpose.

expectations about what respondents would answer played an important role in the coding of ambiguous answers and probing. Specifically, some interviewers judged individual respondents on the basis of their characteristics or previous answers. Furthermore, respondents' answers were evaluated on the basis of interviewers' expectations about what the majority of people would answer. Expectations regarding typical life course trajectories therefore might influence how interviewers lead respondents through the questionnaire. Unexpected patterns might baffle interviewers, which in turn might embarrass respondents or make them unsure about what is expected from them. Atypical trajectories are hence expected to be more problematic and vulnerable to response error than typical trajectories. As Hyman and many researchers after him have shown, interviewer effects are most likely to occur when respondents are uncertain about what the 'right' answer to a question might be.[44]

Time elapsed since last change (in months): Job changes, transfers from one school to another, or transitions between life domains are often accompanied by the preparation of resumes and the review of documents related to one's past. The 'freshness' of biographical memory is especially important for the retrieval of dates. Time elapsed since last change serves as a proxy for the presence of biographical dates in the analysis of visible errors. Unfortunately, in the model for invisible errors the interpretation is less clear. Since number of jobs and events is controlled for, an increase in time since the last transition can mean various things in terms of the effect on the count of invisible errors (see below).

A rather confusing aspect of the design is that the same variable may tap different concepts across the two error count models. Take, for example, the number of job spells before data edition: in the model for invisible errors, this variable is used as a control variable with the hypothesis that the

---

[44]The percentages generated from life course patterns' frequency failed to produce a continuous scale. There was a huge gap, for example, between the classical and most frequent pattern (school-vocational training-work) with 35% and the second-most frequent pattern with 14.5%. I recoded percentages into dummy variables and tested the behavior of these dummies. The recode used empirical gaps in the distribution of percentages as guideline for the creation of these dummy variables. It turned out that, for example, the difference in effects on data accuracy between patterns with 0-1% and patterns with 2-3% was similar to the difference between 14.5% and 35.5%. Hence, the percentage score did not correspond to an interval variable in its effects on data quality. Therefore, the percentages were recoded into an ordinal scale ranging from 1-7 with 7 indicating the most frequent pattern. Arguably, this may force the effect, if any, to be linear; ideally, one would like a complete distribution of life course patterns to test the nature of the continuous effect. Given the data, I cannot do better than inducing a scale on the basis of the observed discontinuous distribution in order to avoid a specification error by introducing the raw percentage scores.

more jobs were already recorded, the smaller the probability that one had been omitted, and a possible curvilinearity to the effect that a lot of jobs already recorded might increase the propensity to omit one, other things being equal.

In the model for visible errors, however, additional job spells might simply indicate a qualitative difference between jobs and other events with respect to their effect on data quality since volume and number of events is controlled for. In order to compare the effect of variables which tap the same concept across equations, I specified both models with exactly the same variables. This might lead to an 'overspecification' and to a certain inflation of parameters but in any case leads me to include variables for which I do not have an a priori hypothesis.

Table III: Differences in Interviewer Performance

workload rank order	mean of visible errors (std.)	rank (visible errors)	% with visible errors	rank (invisible errors)
5	8.73 (6.82)	1	10.64	9
11	9.15 (6.21)	2	11.91	10
21	9.21 (4.29)	3	10.53	8
15	10.90 (5.33)	4	25.00	22
1	11.07 (6.38)	5	14.75	12
16	11.23 (5.52)	6	20.76	17
7	11.41 (6.42)	7	18.52	15
4	11.44 (6.49)	8	22.07	21
12	11.46 (5.67)	9	6.76	2
20	11.68 (6.21)	10	3.57	1
18	11.85 (6.36)	11	31.71	24
13	12.20 (6.79)	12	21.21	18
19	12.22 (7.36)	13	21.95	20
10	12.33 (6.12)	14	8.60	5
23	12.36 (6.28)	15	9.09	6
17	12.42 (6.99)	16	8.00	4
3	12.48 (6.51)	17	19.39	16
14	12.86 (6.53)	18	15.63	14
6	13.04 (6.99)	19	12.41	11
22	13.08 (6.98)	20	7.69	3
8	13.11 (7.24)	21	15.00	13
10	13.19 (8.27)	22	26.88	23
23	13.64 (6.53)	23	9.09	7
2	19.23 (9.91)	24	21.67	19
total	12.33 (7.36)		17.00%	

Note: Differences between interviewers are significant at $p \leq .05$ for all variables. Rankorder for interviewer workload is coded from the highest workload (1) to the smallest (24). All other rankorders are counted from the smallest to the highest value.

IV. RESULTS

4.1. Descriptive Analysis

Before I turn to the multivariate analyses, I present some descriptive material regarding between-interviewer differences in performance related variables. Table III shows huge differences between interviewers in average amount of visible errors, ranging from 8.73 to 19.23, as well as in standard deviations. Interviewers' means on invisible errors range from an average of 3.57% to 31.75% of cases with at least one invisible error.[45]

Table III is ordered by increasing average visible errors; by comparing rank orders in performance across rows, it can be seen that an interviewer with a low average number of visible errors does not usually produce also a low number of invisible errors. For example, the interviewer ranked 11th with regard to visible errors, has the highest percentage of cases with invisible errors. The interviewer with the fewest visible errors is in the middle of the distribution of invisible errors (rank 9), while the two interviewers with the lowest percentage of cases with invisible errors are in the middle of the distribution of average visible errors (rank 9 and 10). On the other hand, the interviewer with the highest average of visible errors is also high on invisible errors (rank 19), and the interviewer with rank 22 on visible errors occupies rank 23 on invisible errors. The Pearson correlation of interviewers' average visible errors with percent cases with invisible errors is r=.328 (P=.12 [under H_0: rho=0]), and drops to r=.185 (P=.40) when the interviewer who scored highest on visible errors is dropped.

Figure 4-7 show field trends in interviewer performance. For the plot of visible errors, I used the rate, namely the ratio of number of visible errors with number of events to avoid confounding effects. Similarly, trends in interview time are shown by plotting the average speed (number of total records per minute of interview time). All observations were pooled by field month and average rate of errors, proportion of cases with jobs added, percent of cases with tape recordings, as well as average interview speed were calculated. Figure 4 shows a clear increase of the rate of visible errors over time. However, Figure 5 does not show a similar time trend in the proportion of cases with invisible errors. There does not seem to be any discernible time trend at all. Average monthly speed,

[45]Similarly, the percentage of cases with tape recordings varies dramatically between interviewer, ranging from 42% to 91%. The average duration of interviews ranges from 52.78 to 109.06 minutes.

Figure 4: Time Trend in Average Monthly Rate of Visible Errors (per event)

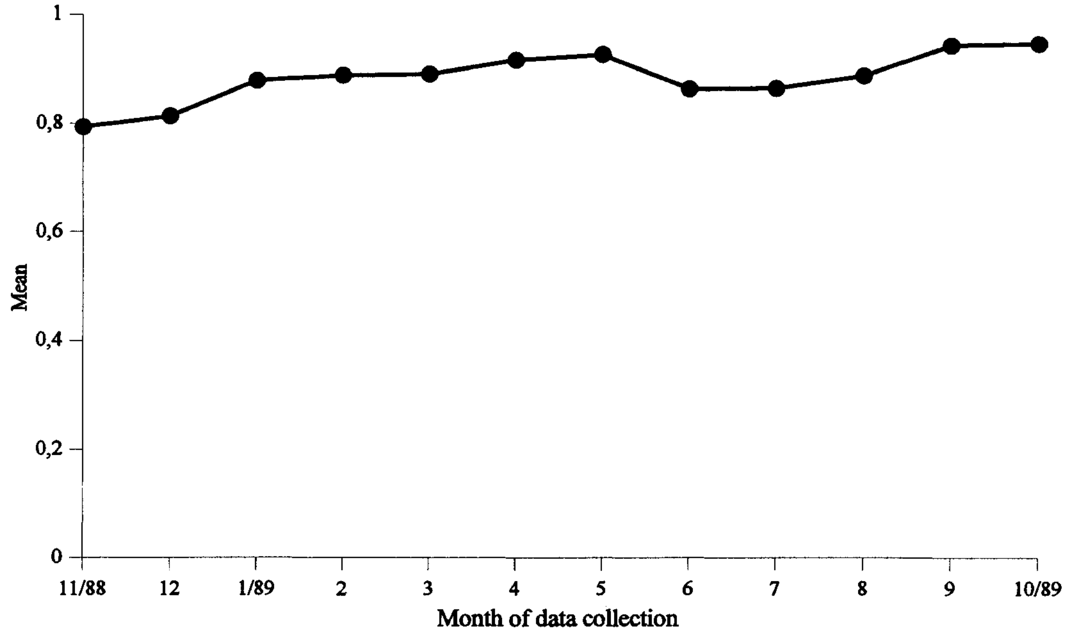


Figure 5: Time Trend in Average Monthly Proportion of Cases w. Invisible Errors

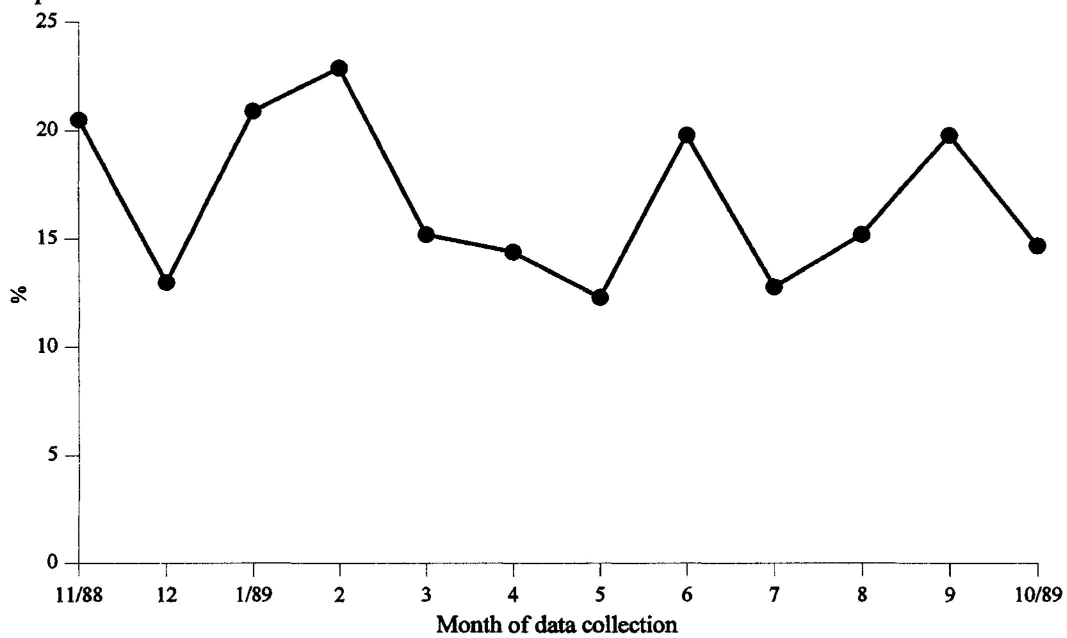


Figure 6: Time Trend in Average Monthly Proportion of Cases w. Tape Recording

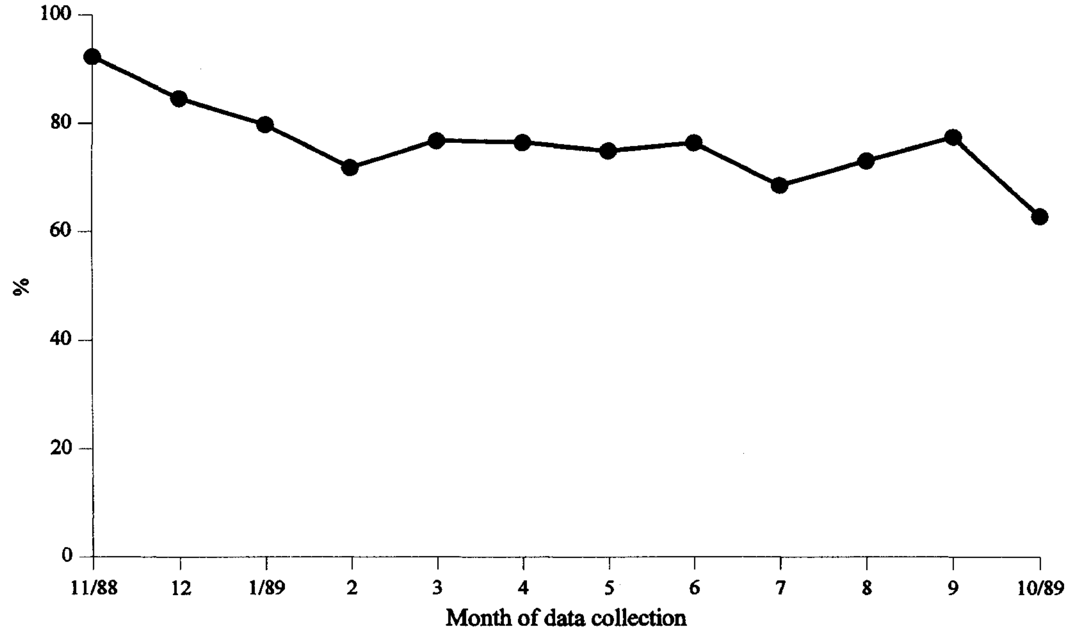
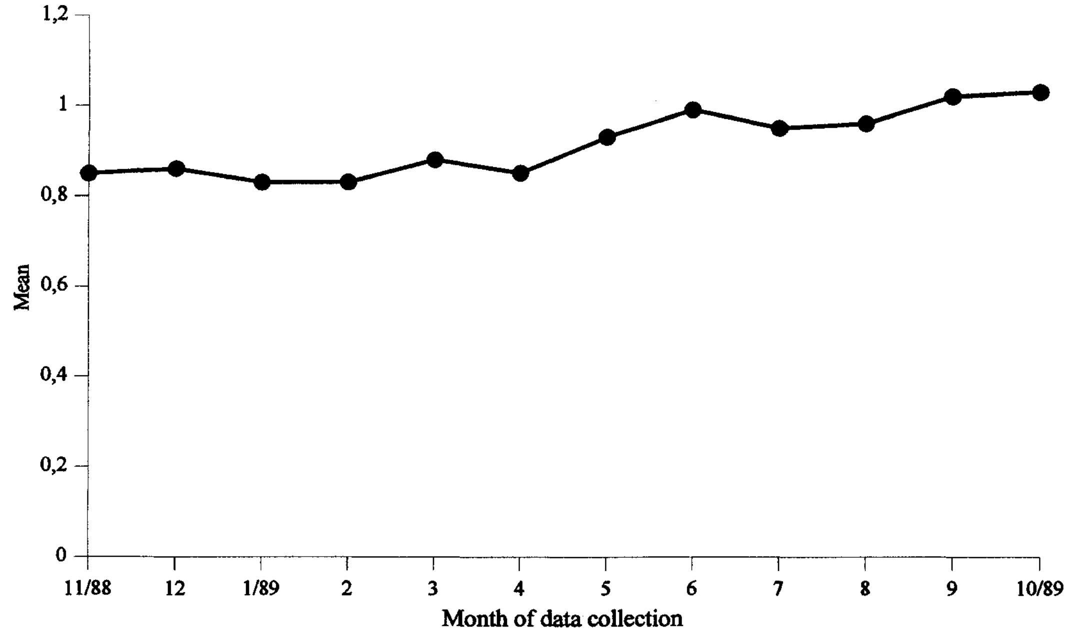


Figure 7: Time Trend in Average Monthly Speed (Records/Minute)



on the other hand, shows a clear increase over time (Figure 7), while the proportion of cases with tape recording declines (Figure 6). Although these trends are confounded with fluctuations in interviewer assignments, they indicate that there may be time trends in performance. Furthermore, they indicate that possible selection effects due to termination of some interviewers' employment do not seem to increase data quality.

The multivariate analysis presented in the next section allows a closer look at these trends while controlling for potentially confounding factors. Before I turn to the interpretation, I report various specification problems for the error count models, and provide some comments regarding the differences between Poisson and OLS models of the logged count variables. After the analysis of outliers and influential cases, I turn to the interpretation of the baseline models and the comparison of effects across models, addressing the again the issue of multidimensionality.

4.2. Multivariate Analysis

4.2.1. Methods and Specification

Both types of errors represent random counts, and previous research has shown that Poisson regression is appropriate for the analysis of these variables (e.g. King 1988). As a log-linear model, Poisson regression differs from ordinary least square regression (OLS) in regard to the distributional assumptions. Limited dependent variables inherently violate the assumption of homoscedasticity of the disturbance. As the mean of the probability distribution gets closer to the lower bound of 0, the variance must become smaller. Furthermore, OLS may yield negative predictions and assumes a <u>linear monotonic</u> effect of X on Y. Taking the log of the count variable (L-OLS) may correct for heteroscedasticity in OLS; however, King (1988) shows that when adding different small constants to Y to avoid the problem of $\ln(0)=-\infty$, size of coefficients and standard errors depend on the magnitude of the added value. He concludes that even logged OLS estimators are biased, inconsistent, and inefficient (p. 847).

The unrealistic assumption of least squares regression of a linear monotonic effect of the explanatory variables on the dependent variables is relaxed in the Poisson model. The Poisson model is based on the assumption that the impact of the explanatory variables depends on the magnitude of the values of the dependent variable. Specifically, the expected impact of a fixed change of x on y_i increases when y_i is already further away from zero, its lower bound. In other words, the "effort" that it would take to move y_i from 0 to 1 should be proportionally greater than the effort required to move

y_i from, say, 20 to 21 (King 1988, p. 842). Under the two assumptions that (1) the observed events within observation i are independent of previous events, and that the probability of an event occurring will be constant within the observation i and (2) $Cov(y_i,y_j)=0$ for $i \neq j$ (no autocorrelation), the Poisson model is characterized by the equality

$$E(Y_i)=VAR(Y_i)=\mu_i.$$

Restricting the expected value to be positive, the equation can be written as

$$E(y_i|X)=\exp(\mathbf{X}_i\boldsymbol{\beta})$$

and

$$P(Y=y_i)=e^{\mu_i}(\mu_i)^{y_i}/y_i! \quad \text{for } y_i=0,1,2,\ldots \text{ and } i=1,2,\ldots N.$$ [46]

However, if the assumption of independence and homogeneity does not hold, the variance will not be equal to the mean. In this situation, Poisson coefficient estimates will still be consistent but inefficient. In the model for visible errors, a contagion process within observations is plausible. Specifically, once the interview gets 'off the track' due to some confusion of interviewer or respondent, dates on subsequent events may be affected.[47] On the other hand, for both models unmeasured inter-respondent heterogeneity may generate overdispersion, e.g. in terms of the capacity or willingness to 'analytically order' the flow of life course events according to the demands of the interview. Therefore, I will test for overdispersion in both models and estimate a negative binomial

---

[46]The log-likelihood functions are described in detail in King (1989), and Cameron & Trivedi (1986).

[47]Various other processes in the interaction may be responsible for overdispersion. Respondents sometimes had problems to remember exact dates especially in the early migration history and interviewers would just assume that this difficulty would apply also to later events and type in missing for months without asking or without motivating respondents to make an effort to retrieve the information.

model (King 1989, Cameron & Trivedi 1986, McCullagh & Neder 1983).[48] Here, the variance of Y is assumed to be greater than the mean and the model is specified as

$$\ln\mu_i = \mathbf{X}_i\mathbf{\beta} + \epsilon \quad \text{where } \exp(\epsilon) \approx \Gamma(1,\alpha)$$

(Greene, 1990). The model has an additional parameter, α, such that

$$VAR(y_i) = E(y_i)\{1 + \alpha E(y_i)\}.$$

The resulting probability distribution is

$$P(Y=y_i|\epsilon) = e^{-\mu_i \exp(\epsilon)}\mu_i^{y_i}/y_i! \quad \text{(Greene 1990).}$$

The interpretation of the estimated coefficients in Poisson regression and negative binomial models is analogous to OLS where $E(\ln Y_i)=\mathbf{X}_i\mathbf{\beta}$ but the effect of explanatory variables on $P(Y=y_i)$ is non-linear and depends on the magnitude of y_i in the Poisson and on y_i and the dispersion parameter α in the negative binomial (McCullagh & Neder 1983, Greene 1990). The Poisson and negative binomial program in LIMDEP was used to estimate the models for error counts (Greene 1990).[49]

In the multivariate analysis, interview time, volume, number of events in the event sequence, and time since last transition were logged to correct for skewness. Table IV reviews and summarizes hypotheses (see section 3.4.2 for explanations).

---

[48]Cameron & Trivedi (1986) mention that the error term in the negative binomial could reflect also omitted exogenous variables (as in the case of unobserved heterogeneity). A non-significant estimated dispersion parameter would could thus be interpreted as test for correct specification.

[49]The LIMDEP default algorithm (D/F/P) posed convergence problems for the negative binomial model. The models presented here were estimated with the Newton algorithm and converged after around 15 iterations when using the Poisson coefficients as starting values for the negative binomial; tests with different starting values and yet another algorithm yielded identical estimates up to the third or fourth decimal place depending on the size of the coefficient.

Table IV: Determinants of Data Quality: Hypotheses

	visible errors	invisible errors
Interview characteristics and volume		
Number of events	+	-
Number of job spells	-	-
Interview volume	+	+
Interview time	+/-	+/-
Respondents' characteristics		
Gender	?	?
Living with partner, presence of small children	+	+
Age	+	+
Schooling	-	-
Life course complexity	+	+
Relative frequency of life course patterns	-	-
Time since last change	+	?
Individual interviewers	?	?
Interviewing		
Tape recording	-	?
Interruptions	+	+
Productivity	?	?
Number of interviewers in shift	+	+
Experience:		
intended learning processes	-	-
unintended learning processes	-	+

Table V reports the models for both visible and invisible errors. The main problem was to find a model to demonstrate the effects of interview time and volume which worked equally well for both types of error. When introducing volume in the equation, the dispersion parameter a becomes insignificant in the model for invisible errors (model not shown, a=.1707, s.e.=.1078, t=1.584). When volume is omitted, the estimation yields a significant dispersion parameter (a=.125, t=2.30) which may be due to under-specification (Cameron & Trivedi 1986). In the following, I report always the results of the Poisson model for invisible errors and the negative binomial for visible errors. To compare nested models, I performed a likelihood ratio test.[50] It turns out that a specification containing an interaction term between time and volume yielded a slightly better fit in the model for invisible errors (LR=3.84 with 1 df) although it does not significantly improve the fit of the model of

[50]The formula is given by $LR=2[\ln L(\hat{\Theta})-\ln L(\hat{\Theta}_0)]$ which asymptotically follows a χ-square distribution with k degrees of freedom (k=number of restrictions). The likelihood ratio test reported in the tables, unless indicated otherwise, compares the model presented in the table with a model containing the intercept only for Poisson regression models, or the intercept and the dispersion parameter for the negative binomial model.

visible errors (LR=2.2).[51] In order to compare the effects across models the interaction term is retained in both models.

Similarly, when testing for the effect of schooling it turns out that using a set of dummy variables for the degree categories[52] instead of the ordinal scale improves the model fit for visible errors (LR=12.3 with 3 df) but not for invisible errors (LR=2.9 with 3 df). Examining the coefficients for the 4 dummy variables shows that visible errors are expected to decline with increasing education, but the effect is not quite linear. Duration of schooling was not significant in any of the models (LR=1.2 for visible errors and 3.2 for invisible errors) when controlling for degree and is therefore dropped. Subsequently introducing squared terms for interviewer experience, number of events, and number of jobs to test for curvilinearity does not improve the model fit, and the terms are not significant.

4.2.2. Comparison of Event Count Models with OLS and Diagnostics

Diagnostics were performed on the baseline model, using OLS regression models of the logged counts to identify potential outliers and influential cases.[53] The model was then refitted for both error types when deleting cases exceeding the cut-off values on one or more of the following statistics: studentized residuals, Dfitts, Cook's D, and leverage values from the HAT matrix (Bollen & Jackman 1990). For the Poisson model, diagnostics are less well developed and readily available than for OLS regression. Although the estimates in the L-OLS models are known to be off, the results at least for the model of visible errors are relatively similar, so that this procedure may be justified.

Appendix A reports estimates of coefficients and standard errors for visible errors. Results of the negative binomial are repeated in the first two columns for convenience. Both coefficients and

---

[51]I also tested for a curvilinear effect of time. The squared term is not significant in the model for visible errors and the overall model fit is slightly lower than in the interaction model (Loglikelihood=-5545.8 as compared to -5543.9 in the model presented in table V). However, in the model for invisible errors, the squared term was marginally significant (t=1.699) while the model fit is almost identical to that of the interaction model (Loglikelihood=-951.28 as compared to -951.35 in the model reported in table V). Introducing both terms in the model caused convergence problems in the negative binomial and did not improve the fit of the Poisson. Since the two specifications seemed equivalent in terms of fit I decided to use the interaction model for easier interpretation and the test of other interactions with time.

[52]Using the degree measure for education yields a better model fit than duration of education (see section 3.4.2).

[53]The log was formed by adding 1 to the count variables.

standard errors are very similar in expected size and directions of effects with the exception of the coefficients for logged time, volume, and the interaction effect time*volume. This difference is probably due to the high collinearity of the variables (the variance inflation in the L-OLS model is estimated as 162, 131, and 467 respectively for the three parameters). A correct specification of the functional form in the negative binomial may account for a more efficient use of the information. The coefficients for tape recording and weekly productivity are slightly larger and significant in the L-OLS model, while they only approach borderline significance in the negative binomial (t=-1.595 and 1.721 respectively). Similarly, the coefficients for the dummy variables flagging interruptions in the interview are much smaller in the L-OLS model and not significant (t=1.578 and 1.489). In sum, both with respect to estimates and substantive conclusions the two models are fairly similar with the exception of interview interruptions.

A comparison between Poisson and L-OLS of the model for invisible errors, reported in Appendix B, shows that here the estimates are dramatically different in size.[54] Generally, the L-OLS coefficients have the same sign and significance as in the Poisson model but are much smaller. In terms of the substantive conclusions drawn in the next sections, tape recording and weekly productivity are significant in the Poisson but not in L-OLS model, and interviewer experience is significant only in the L-OLS model. These differences, however, may be expected for variables with borderline significance. The large differences in size, on the other hand, indicate that the specification error in the L-OLS model is more aggravating for the model for invisible errors than for visible errors. The latter variable may approximate more closely a continuous normal distribution as assumed in OLS, whereas the bias introduced by adding a constant to the raw score in order to form the log may exacerbate problems in the L-OLS model for invisible errors since most cases score zero on this variable to begin with.

For the identification of outliers and influential cases, I used cut-offs as described in Bollen & Jackman (1990). For the model of visible errors, 8 cases exceeded the high cut-off on one or more statistics, and 107 exceeded the low cut-off on two or more. Refitting the model yielded virtually the same results when 'high' outliers were removed; the model fit is identical with 1912.8 compared to

[54]Diagnostics on the model for invisible errors were performed including the log of visible errors as explanatory variable. The relationship between visible and invisible errors is important for conclusions about field organization and the nature of data quality. Therefore, I retained visible errors in the equation for invisible errors for both the regression diagnostics and the analysis of interaction effects to make sure that the absence of an effect holds across different specifications and samples. See below in the section 4.5.

1912.4 in the full sample. When observations exceeding the low cut-offs were removed, the model fit improved (LR=2238.8). The interaction term between time and volume becomes significant and somewhat larger than in the full sample (ß=-.1563, t=-2.104). Also, the coefficient for productivity becomes significant, and the dispersion parameter (α=.0108) is reduced to a third of its size in the full sample. Possibly, with the removal of outliers, either unmeasured heterogeneity among respondents or processes of contagion across events are reduced.

Comparing the means on these variables with the full sample shows that in the outlier sample interviews took on the average about 1 minute longer while the volume was slightly smaller with an average of 56 records compared to 58 records in the full sample.

Applying the same procedure to invisible errors yielded 14 'high' outliers and 133 'low' outliers. The most interesting result is that 55 cases were identified as outliers both with respect to visible and invisible errors. The mean of invisible errors in the outlier sample is much higher with 1.98, as is time with 75 minutes (compared to 67 in the full sample) and volume with 66. Also, outliers tend to be disproportionally male. Furthermore, in the outlier sample for invisible errors the average of visible errors is higher than in the full sample (14.04 compared to 12.33). This finding shows that while the two indicators of data quality may tap distinct processes, at the extremes they move together. This is supported by the fact that in 65% of the 133 outliers on invisible errors call-backs were made, as compared to 25% in the full sample! It may hence be justified to classify these cases as 'difficult' not only in terms of the models presented here, but in every aspect of data collection.

Deleting the 14 'high' outliers and refitting the model for invisible errors yielded essentially the same results as presented in Table V. Refitting the model to a sample without the 'low' outliers (N=1854), however, reduces the model fit (LR=333.6). The differences between estimates for the two samples are considerable and reported in Appendix C. In the remaining sample, the mean of invisible errors is reduced by half. Compared to the full sample, in the refitted model the coefficient for time changes size and sign. The interaction term between volume and time loses significance (t=-1.381), while the main effect of volume becomes larger and significant. Apparently, the L-OLS model used to identify these outliers has a hard time explaining the occurrence of invisible errors, so that cases with such errors are disproportionally picked up as outliers. It is perhaps questionable how good a strategy for diagnostics this procedure is; it yielded, however, an interesting sample of cases to which I will refer below as 'difficult' cases. Finally, I deleted 14 cases which were flagged as extremely problematic during data editing. For two, data collection stretched over more than two months; others were noted because respondents had great difficulties to remember event dates or were just very uncooperative. Most of these case (N=10) were also picked up by the diagnostics, three with respect to both models.[55] The results correspond to those yielded by removing outliers exceeding the high cut-offs for both models.

4.3. Interpretation

According to McCullagh and Neder (1983), the 'elasticity' interpretation in double-log OLS models holds also for Poisson and negative binomial models. Thus, for logged independent variables, a 1% change in X leads to ß% change in the expected mean of Y. Dummy variables can be interpreted in terms of $\hat{g}$=exp(ß)-1 with an expected 100$\hat{g}$% difference in Y between the two categories flagged by the dummy variable (McCullagh & Neder 1983, p. 139). For unlogged independent variables, a 1-unit change in X is expected to lead to an exp(ß)% change in Y.

4.3.1. Visible Errors

Table V, column 1, reports the estimated coefficients for visible errors. The model fits the data considerably better than a model with the intercept and the dispersion parameter only (LR=1912.4 with 41 df). The dispersion parameter α is estimated as .0407, highly significant (t=12.93) and stable across different specifications of the model. Both intra-respondent, inter-event contagion and unobserved inter-respondent heterogeneity are plausible mechanisms which may generate overdispersion.

Interview. With increasing number of events, the expected number of visible errors increases disproportionally: given a 1% increase in events, the average number of errors is expected to increase by 1.5%, other things being equal. When controlling for volume and the other variables, with increasing time, the number of errors was hypothesized to increase, indicating difficulties with the questionnaire. This hypothesis is supported by the model with an expected percentage change in errors of .7 (however, this coefficient corresponds to the expected change when volume is equal to zero in the context of this model). The same model without the interaction term yields a coefficient of .1760 (t=3.838) for time and -.3081 (t=-4.787) for volume (model not shown).

---

[55]The remaining four cases are problematic with respect to the data on family members or other problems which my indicators for data quality could not pick up.

Table V: Baseline Model for Visible and Invisible Errors. Negative Binomial and Poisson Estimates.

Variable	Visible errors coefficients	s.e.	Invisible errors coefficients	s.e.
Constant	-3.0462*	1.3876	2.7618	8.5263
Interview				
Log of time	.6577*	.3278	-5.1385*	2.0735
Volume	.1894	.3430	1.4849	2.0217
Log time*volume	-.1178	.0795	.9760*	.4887
Log # of events	1.4844**	.0604	-2.7162**	.2724
# of job spells	-.0354**	.0061	-.3370**	.0410
Interviewing				
Tape recording	-.0328	.0206	.2518*	.1248
Experience	.0011**	.0003	-.0028	.0015
Productivity	.0124	.0072	-.0794*	.0364
# interviewers in shift	-.0026	.0038	.0118	.0213
Two sessions	.0668**	.0192	.1217	.1195
Three + sessions	.1066*	.0509	.5945*	.2690
Respondent				
Female	.0068	.0177	-.8137**	.1077
Age	.0121**	.0036	.1082**	.0200
Time since transition	.0081	.0127	-.3557**	.0605
Trajectories:				
Frequency	-.0058	.0055	.0517	.0369
Complexity	-.0039	.0081	.1284**	.0494
Schooling:				
Hauptschule	-.0958*	.0452	-.3158	.2000
Middle School	-.2418**	.0469	-.3224	.2088
Fachhochschulreife	-.2929**	.0531	-.7614**	.2499
Abitur	-.3366**	.0477	-.8333**	.2161
Living with partner	-.0036	.0208	-.1837	.1207
# of small children	-.0214	.0137	-.0421	.0669
Individual Interviewers				
1	-.0079	.0715	.7684	.4534
2	-.1975**	.0578	-.2853	.4770
3	.1349*	.0618	.2725	.4377
4	-.3931**	.0541	-.2919	.4384
5	.3396**	.0553	1.2518**	.4007
6	-.0252	.0869	-.2943	.5786
7	-.1287*	.0583	.5290	.4130
8	-.0158	.0578	.5263	.4114
9	-.1247*	.0544	.4599	.4110
10	.2116**	.0576	.9326*	.4087
11	.0002	.0726	.9917*	.4212
12	-.0116	.0568	.2257	.4112
13	.0016	.0691	-.6531	.5757
14	.2750**	.0686	.9471*	.4469
15	.1221	.0681	.6489	.4682
16	.0488	.0637	.3225	.4340
17	.0396	.0540	.3090	.4048
18	.0306	.0771	.7425	.4299
19	.0407	.0596	.8289	.4490
Dispersion parameter α	.0407**	.0032	--	
Likelihood ratio	1912.4 with 41 df		678.4 with 41 df	

*$p \leq .05$ **$p \leq .01$

Hence, visible errors increase with increasing time when controlling for volume, while increasing volume leads to a decrease in errors when controlling for time. The model offers only inconclusive support for an interaction between interview volume and time. The interaction term is not significant even when performing a one-tailed test, but it is not too far from significance (t=-1.483). When deleting the outliers identified with L-OLS, however, it becomes significant, indicating that, at least for a reduced sample, the positive effect of time on visible errors might be attenuated as volume increases, indicating an unhampered flow of information, other things being equal. Similarly, the positive effect of volume on errors may be attenuated with increasing levels of time, indicating that interviewers and respondents produce better data when they take more time to produce the same amount of data.

Each additional job spell is expected to decrease the expected mean of visible errors by about 1% net of the other variables. The coefficient is highly significant (t=-5.759) and can be interpreted as showing the higher salience of labor force participation and hence, enhanced data quality as the number (i.e. proportion, since number of events is controlled for) of jobs in the event sequence increases.

Interviewing. A variable of great interest in this paper is the effect of interviewer experience on data quality. Each additional completed interview is expected to increase the average visible errors by 1% net of other factors in the model (t=3.873). This result runs counter to expectations in that apparently interviewers do not produce better data quality with growing experience. Unintended learning processes, if any, seemed to have produced a 'sloppier' interviewing style since this finding cannot be explained 'away' by any respondent characteristics--the predicted increase therefore seems to stem from interviewer behavior such as skipping questions related to dates in the event sequence. The time trend in visible errors illustrated in Figure 4 is thus confirmed by the multivariate analysis.

Weekly productivity has a positive effect with borderline significance (t=1.721) on visible errors. Removing the outliers from the sample yields a significant effect of productivity with about the same size (t=1.970). Thus, the more interviews in a given week an interviewer conducts, the more visible errors on the average the interviews will have, consistent with the image of a burn-out effect. This effect may be offset when difficult cases are encountered (namely, the outliers): in the latter situation, routine and easiness of handling respondents, computer, and questionnaire may enhance data quality. This, then, might result in the larger standard error of the productivity coefficient in the full sample.

The expected mean of visible errors for interviews with tape recording[56] is on the average 3.2% lower than that of cases without tapes, net of other variables. Considering the cost, this increase in data quality seems rather small, and the coefficient is not significant (t=-1.595). With respect to visible errors, one might conclude that tape recordings are less efficient as a field control mechanism and, in the present survey, were important mainly as data source for editing. The positive, if small, effect of tape recording on interview time shown in Appendix D, however, supports the hypothesis that interviewers develop negatively sanctioned strategies to save time when the risk of exposure is low. The hypothesized negative effect of tape recordings on visible errors seems therefore to operate mainly through taking more time to do the interviewing.

The estimated effect of the number of interviewers working in a shift has an unexpected negative sign but is small and insignificant (t=-.672). While the working place environment seems thus less important for data quality, there is strong support for the hypothesis that interruptions of interviews lead to an increase in the error count. Interviews conducted in two sessions are expected to have on the average 7% more visible errors, two or more interruptions lead to an expected increase of 11%, net of other variables.

Respondents' characteristics. The effect of education on visible errors is pronounced with a predicted 29% reduction in the expected mean of visible errors for respondents with Abitur as compared to respondents who never acquired a formal certificate for school, holding constant other variables. For people who graduated from the Hauptschule, the expected decrease in visible errors is markedly lower with 9%, and people with middle school degree or Fachhochschulreife lay in between with 21% and 25% respectively. The effect of education on visible errors is hence monotonic but not linear, with middle school degree as an apparent threshold.

Respondents' age has the expected positive effect on the error count. Each additional year increases the expected mean of errors by about 1%. Since the number of events in the interview is controlled for, this effect may point to the effect of the length of time elapsed between the actual events and the interview and is consistent with the results of Wagner (1989). In order to further

[56]More than 70% of all interviews were taped. As explained above, these tapes are a rich source of information. In the present context, the variable aims at tapping the concept of control over the interviewing style. For almost a year, interviewing continued while the completed interviews were submitted to data editing in Berlin. Data editors could use the tapes to give feed-back to interviewers; thus, any negatively sanctioned behavior like skipping or altering questions had a certain probability to be detected where a tape recording was present.

explore this effect, it would be necessary to estimate the probability of the occurrence of an error in a given "spell" dependent on the time elapsed between the interview and that event.

Time since last life course transition has the expected sign, accounting for a proportional increase of errors as it increases but is not significant. Based on Table V, one might infer that females do not differ from males net of other variables, and that the same is true for persons living with a partner compared to singles. The negative sign of the coefficient of the number of small children living in the household is unexpected although not significant. I found, however, a significant interaction between respondents' gender and the presence of small children in the household in the model for visible errors (ß=-.0566, t=-2.083). When introducing an interaction effect between being female and both number of children under age six and living with a partner in the same household, the model fit improved (LR=12.4) and the interaction with living with a partner was also negative but not significant (ß=-.0631, t=-1.636). Also, the coefficient for female respondents becomes positive and significant (ß=.7847, t=2.727). Hence, compared to male respondents without small children, females without children are expected to have on the average more visible errors, while mothers have less. The coefficient for males with small children was slightly positive and not significant (ß=.0027, t=.200). In short, other things being equal, rather than more distracted, mothers of small children seem to be more organized than other people.

Life course pattern frequency has the expected sign, indicating that frequent and perhaps expected patterns are easier to handle than rare and unexpected patterns. Life course complexity, on the other hand, has the wrong sign. Possibly, its effect is conditional on other variables (see below). Both coefficients are not different from zero in this model and relatively small (translating into a .6% and .4% decrease in the expected mean of errors per unit change respectively). I tested whether the effect of life course complexity is conditional on time. The interaction term is negative (ß=-.0557, t=-2.300) and the main affect of complexity becomes positive (ß=.2328, t=2.215), while the main effect of time remains fairly similar to the baseline model with ß=.7950 and t=2.363. The hypothesis that complexity contributes to visible errors is confirmed. The effect, however, depends on how much time is taken for the interview. That is, the positive effect of complexity on visible errors is attenuated by additional interview time.

Individuals matter. Lastly, for 10 of the 19 dummy variables flagging individual interviewers, the model predicts a higher average error count than for the 5 interviewers who completed less than 30 interviews while for the other 9 interviewers a lower average is predicted. The inclusion of the interviewer dummy variables improves the model fit significantly (LR=408.4 with 19 df). Eight coefficients are significant, some translating into substantial percentage changes in the expected mean of visible errors compared to the baseline category and net of other variables.[57] For example, while the interviews conducted by interviewer No. 4 are expected to have a 32% lower mean of errors compared to the default category, the mean of those conducted by interviewer No. 5 is expected to increase by 40%! In section 4.4. below I will address the question whether these substantial differences are paralleled by inter-individual differences in the effects of performance related variables.

4.3.2. Invisible Errors

In contrast to the previously analyzed visible errors, invisible errors are the operationalization of a "deeper" concept of data quality. Specifically, they should measure the extent to which the concept of a job spell was communicated and interpreted correctly in the interview. That is, the extent to which changes in the job history experienced by respondents such as changes in the actual content of job, firm, local establishment, and position, were reflected in the data. In contrast to invisible errors, they cannot be identified by a set of logical rules as applied in the check program; only a systematic, substantive comparison of all available information can help uncovering these errors. Table V, columns 3 and 4, reports the results for invisible errors. Introducing the co-variates improves the model fit significantly (LR=678.4 with 41 df).

Interview. The more information a case contains, in terms of number of events, and number of job spells, the lower the chances that a job spell had to be added. However, when time is at zero, increasing volume has a positive effect on invisible errors net of other variables. In contrast to the model for visible errors, Table V shows a significant interaction between volume and time on the expected mean of invisible errors. The effect of volume is conditional on interview time; at a minimum value for time, a 1% change in volume leads to an expected 4.6% increase of average invisible errors. At the mean of time, the effect of a 1% change in volume increases to 5.5%, and at time's maximum

---

[57] The other coefficients are not only not significant but also, with one exception, much smaller than the significant interviewer variables. Overall, a comparison of significant with insignificant parameters indicates that where effects are relatively large they are also significant and conclusions are not limited by a combination of a large effect with a large standard error as may be the case when too few cases fall into a group. Choosing another baseline, namely the interviewer who had the highest workload (model reported in Table VI), does not significantly improve the model fit (LR=3.2 with 4 df). Also, the additional coefficients restricted to zero in the model reported in Table V are small and far from significant.

the effect is 6.7%. Hence, the longer the interview takes, the stronger the tendency of interviewers and/or respondents to try to take a short cut, e.g. by omitting or collapsing one or more job spells. Similarly, the more time the actors take for producing information, the lower will the expected mean of invisible errors be; with increasing levels of volume, however, the negative effect of time is attenuated. A 1% change of time leads to an expected 2% reduction of errors at the minimum value of volume. The effect reduces to 1.2% at volume's mean and to .2% at a maximum value for volume. As Appendix C shows, however, the interaction effect loses significance when outliers are deleted, and the coefficient of time becomes positive, similar to the effects shown in the model for visible errors. Hence, the dynamic relationship between time and volume described here is driven by the inclusion of these cases which amount to 6.7% of the full sample, and are above average with respect to both time and volume. Refitting the model when omitting the interaction effect yields a significant negative coefficient for time and a significant positive effect for volume, that is, the opposite combination than in the model for visible errors.

Interviewing. In regard to interviewer experience, in this model we find the opposite effect than in the model for visible errors. Here, each additional interview is expected to decrease the mean of invisible errors by about 1% (ß=.0028, t=-1.893, p=.058, borderline significance). Similarly, weekly productivity is significantly and negatively related to invisible errors (t=-2.181). Both effects indicate positive and 'intended' learning processes in that interviewers get better at 'getting the story across' the more and the longer they work. These effects are somewhat different when the difficult cases identified in the diagnostics section are removed--the negative effect of experience becomes larger (ß=-.0042, t=-2.010) and significant, while the effect of weekly productivity becomes somewhat smaller and loses significance (see Appendix C). This reversal in the significance brings out the difference between long-term and short-term learning effects: day-to-day routinization pays more when a difficult case is encountered, so to speak.

Interruptions of the interview tend to increase invisible errors. Especially interviews with two or more interruptions (N=45) are expected to have a 80% increase in the mean of invisible errors (t=2.210) compared to interviews which were conducted in one session, net of other variables. In the sample without outliers (Appendix C), this effect is reduced almost by half and far from significant. The reduced sample contains only 34 of the 45 cases with more than two interruptions, and apparently these 10 cases account for the significant effect in the model for the full sample.

The number of interviewers per shift has a small positive insignificant effect which confirms the relative unimportance of this aspect of the work place environment found in the model for visible errors. The positive effect of tape recording on the number of job spells added (t=2.017) is explained in terms of the editing process. Apparently, in spite of the call-backs which were more frequent for interviews without tapes, more information was more readily available for taped interviews.

Appendix F shows that interviewer experience and tape recording had a positive effect on the number of job spells before data editing. Unfortunately, while eliminating interviewing and interviewer effects on the number of job spells--all effects are reduced in size and insignificant in the model for jobs spells after data editing--during the data editing "editing effects" may have been introduced.

Respondents' characteristics. Interviews with female respondents are expected to have on the average 56% less invisible errors than interviews with males, other things being equal (t=-7.556). This somewhat unexpected finding in any case rejects the hypothesis that an allegedly lower salience of labor force participation for women would yield lower data quality with respect to jobs. If salience is a factor at all, it operates in an unexpected direction. Alternatively, women may tend to be the 'keepers' of biographical facts, or the effect may be due to discriminatory data editing.[58] Interactions of gender with the number of small children living in the household and living with partner are not significant in this model.

For respondents' age, each additional year is expected to lead to an expected increase in the mean invisible errors of 1.1% (t=5.403). The highly significant (t=-5.877) while small -.7% expected decline in invisible errors for a 1% change in the time since last transition is not to explain with retrieval processes--if this variable indicates increasing difficulty to recall events more removed in time, its coefficient should be positive as in the model for visible errors. Possibly, the variable indicates in the context of this model a certain stability in respondents' life courses, which might reduce, so to speak, the 'opportunity' to omit a job spell compared to other, more instable life courses.

A one unit change in life course complexity is expected to increase the mean of invisible errors by 1.1%. In contrast to the model for visible errors, an interaction of interview time and complexity was not significant. Education has a negative impact on invisible errors, especially for people with more than a middle school degree.

---

[58]Appendix F shows that women currently living with a partner are expected to have on the average less job spells than men and single women. This is a commonplace finding within the German context and it is impossible to say how much the expectations of all actors participating in the survey contribute to the negative coefficient.

Individuals matter. Only four interviewer dummy variables are significant in this model but it is noteworthy that compared to the five interviewers who completed under 30 cases they are all positive and large. Adding the interviewer dummy variables improves the model fit (LR=79.8 with 19 df). Three of these interviewers have also positive and significant effects on visible errors. When choosing the interviewer with the highest work load as comparison category, four interviewers emerge with large significant and negative coefficients. This finding points to the importance of improving field control; neither additional training, nor feed-back from the editing group nor even supervision during the interviewing helped to prevent a situation in which some of the most active interviewers in the field produced consistently below-standard data quality, at least in terms of the two measures employed in this analysis.

4.4. How Much Do Individuals Matter?

A comparison of the results from descriptive and multivariate analysis in regard to the difference between the two types of error leads to the question whether the different effects of experience and productivity on the two types of error hold for all interviewers alike. Therefore, I introduced interaction terms with interviewers and both experience and productivity. To avoid problems with collinearity and convergence, I introduced interaction terms in groups of four or five.[59] In the model for visible errors, interactions of interviewers with productivity yielded no better model fit except for two interviewers (No. 15 and 16 in Table V), for whom the effects were positive and significant. For the model of invisible errors, on the other hand, where the effect of productivity was negative, for two interviewers the effect was positive and significant (No. 6 and 13). I found only one significant interaction effect of interviewers and experience on invisible errors which was positive, in contrast to the overall effect, which is expected to be negative. For three interviewers the effect of experience on visible errors was negative and significant, for another three positive.

When using the predicted main effect of individuals or overall workload as classification criteria, a pattern giving meaning to these interaction effects did not emerge (e.g. that the effects of experience and productivity were systematically different for 'good' interviewers or for 'bad' interview-

---

[59]Due to limitations of LIMDEP, I could not introduce all interaction terms at once in the model. When doing so in the L-OLS model, the improvement in fit for visible errors is not significant although three interaction terms are significantly different from zero (one with a positive sign, two with a negative sign). In the L-OLS model for invisible errors, multi-collinearity precludes further analysis.

ers). For example, among the three interviewers with a significant negative interaction effect on visible errors with experience, we find the 'best' and the 'worst' interviewer (No. 4 and No. 5 in Table V), while the third has a zero effect on visible errors but a significant positive effect on invisible errors. It is perhaps just as well, since with so many parameters tested in the models, some interactions may come out significant by chance alone.

Hence, I finally ask whether the effects observed in the overall sample are the product of just a few interviewers. When deleting the three interviewers with significant positive interactions with experience from the sample, the effect of experience is reduced to about a third of its original size and loses significance (ß=.0003, t=1.192). Taken together, these three interviewers completed almost a fifth of all interviews (N=389) which explains their influence on the sample estimates. Similarly, when deleting the two interviewers with significant positive interactions with productivity, the coefficient is reduced to .0079 (t=1.028). For invisible errors, no significant interactions pointing in the same direction than the overall effect were found; when deleting the three interviewers with the largest estimated effects, results remain unchanged. To make a long story short, the positive effect of experience and productivity on visible errors is largely driven by a few interviewers, the latter effect being significant only when deleting outliers, as it were. The negative effect of productivity on invisible errors remains unchanged, while the evidence for a negative effect of experience was inconclusive in the first place but is not further undermined.

V. DISCUSSION

5.1 Visible and Invisible Errors Compared

I have argued that data quality in the context of restrospective survey like the present is best seen as a multi-dimensional construct. Comparing effects of independent variables across models thus seems a convenient starting point for the discussion of results. This comparison shows important differences and similarities (see Table V).

Interview. The difference of effects of time and volume across the two models points to a complex situation. Interview time was shown to be a proxy for two processes in the interviewing process: on the one hand, it is an indicator of the efficiency of the interview--as such, it indicates confusion and recall difficulties in the event sequence. On the other hand, time may be an indicator of tendencies towards 'output reduction' when interviewers skip questions, when time for complex memory retrieval processes is not taken or given, and when job spells are collapsed. My results show that, for the full sample, the first meaning of time is prevalent in the model for visible errors (more interview time indicates less efficiency), while in the model for invisible errors, the second aspect of time seems to drive the effect (more time indicates more throuroughness). Thus, a better 'flow' of information reduces visible errors, while, for invisible errors, more time simply means more information.

The situation is further complicated by the fact that the interaction of volume and time is significant for invisible errors only in the full sample, while, for visible errors, it is significant only in the sample where outliers are deleted. If it is correct to interpret the interaction of time and volume in the model for invisible errors as sort of a ceiling effect (at high levels of volume, the negative effect of time is reduced), it is not surprising that it loses significance when deleting the longest and 'fullest' interviews. For visible errors, on the other hand, taking more time to produce the same amount of information may be 'effective' only for cases which are not exceedingly difficult in the first place (and hence removed as outliers).

Interviewing. Most importantly, do effects of interviewer experience and weekly productivity operate in different directions? Table V shows that both variables are expected to increase visible errors but reduce invisible errors, other things being equal. The inspection of outliers further supported this conclusion, showing that these effects are not generated by influential cases but are present or even more pronounced also in reduced samples. They are generated, however, in part by influential interviewers, and it is hard to make inferences from these data.

With appropriate caution, I conclude that the analysis yields modest support for a differential impact of experience on the two types of errors, at least for some interviewers. If so, the result is somewhat at odds with the hypotheses listed in Table IV. Neither did experience help avoid both types of errors, nor did interviewers learn to avoid visible errors but reduce output by making invisible errors. In sum, with respect to invisible errors, intended learning effects, if any, may have taken place, while the increase of visible errors may point to a lack of efficient field control.

Table VI: Coefficients for Interviewers when the Interviewer with the Highest Work Load is the Default Category.

	visible errors		invisible errors	
Interviewers	coefficients	s.e.	coefficients	s.e
from Table V:				
1	-.0460	.0741	-.0604	.4122
2	-.2344**	.0600	-1.1180**	.4285
3	.0973	.0652	-.5583	.3873
4	-.4307**	.0563	-1.1246**	.3838
5	.2974**	.0534	.4253	.3206
6	-.0629	.0890	-1.1254*	.5410
7 (baseline)	.0	.0	.0	.0
8	-.1695**	.0569	-.2987	.3352
9	.1012	.1218	-.8396	.7667
10	-.1656**	.0535	-.3680	.3141
11	.1732**	.0593	.1032	.3534
12	-.0398	.0722	.1617	.3526
13	-.0496	.0583	-.6067	.3443
14	-.0367	.0698	-1.4844**	.5317
15	.2352**	.0685	.1212	.3909
16	.0832	.0694	-.1799	.4200
17	-.0673	.0790	-.8131	.7633
18	-.0900	.1211	-.4626	.7642
19	-.0082	.0761	-.0870	.3625
not in Table V:				
20	-.0565	.0583	-.3030	.3448
21	.1174	.1248	-1.1171	1.0424
22	.0077	.1101	-1.0862	1.0391
23	.0098	.0617	-.0573	.3632
24	-.0001	.0556	-.5196	.3406
Likelihood ratio compared to Table V	3.2 with 4 df		.34 with 4 df	

Interviewers. Three interviewers with positive significant effects on visible errors show the same for invisible errors. Looking at Table VI, we can see that, compared to the interviewer who completed most cases, two interviewers have significant negative effects on both types of errors. The rest of the 19 interviewers have significant effects only on one of the errors types, for four the signs of the coefficients are reversed across equations. In sum, interviewer effects on the two indicators of data quality are consistent for some but not all interviewers.

Respondents. Respondents' education reduces errors of both types. I don't quite know what to make of the different thresholds shown in the two models--for visible errors, the largest difference in coefficients was between people graduated Volksschule and middle school, while for invisible errors, the jump was between middle school and Fachhochschulreife.

Life course complexity, namely the fragmentation of chronological sequences into life domains, has a direct positive effect on invisible but not on visible errors. When introducing an interaction of interview time with life course complexity, however, it turned out that the effect of complexity on visible errors is conditional on time. That is, an increase of visible but not invisible errors could be and was avoided by taking more time for recalling and probing.

Age increases both types of errors for the two cohorts considered in this paper. A comparison with older cohorts in the Life History Study and in other retrospective surveys might prove interesting. The same is true for the effects of gender shown above; in regard to invisible errors, female respondents' data are better regardless of living situation, while with respect to visible errors, single women have more errors than men, and mothers less than other people.

Have I shown that data quality is best analyzed as a multidimensional concept? Especially with respect to interviewer effects, experience and productivity, my results are perhaps less conclusive than I would like them to be. On the other hand, with respect to the dynamics of time and volume, as well as with respect to respondents' attributes, I find interesting differences in effects. It is not always be unambiguous what these differences mean; before I proceed to the next section and some remarks about limitations in data and design, I report results of the last analysis carried out in this paper, attempting to trace the relationship between visible and invisible errors.

Appendix C, column 1, reports the coefficients for invisible errors with visible errors added as an explanatory variable (logged to correct for skewness). If visible and invisible errors were generated by basically the same processes, visible errors could be interpreted as a proxy for these processes insofar they are not accurately measured by the other variables in the model. On the other hand, editors may be more likely to add job spells in cases which already showed a lot of errors in the event sequence.

The results from the multivariate analysis confirm that the two types of errors are only weakly related although the most 'difficult cases' tend to be high on both. Clearly, visible errors do not 'cause' invisible errors or indicate some propensity to make errors not captured by the other variables in the model. The effect is far from significant and also very small with an expected increase of .08% in the mean of invisible errors for a 1% change in visible errors (t=.546; LR=.26 with 1 df). Estimates for all other variables are very similar to those reported in the baseline model (see Table V).

5.2. Strength and weaknesses of data and design

Before I proceed with the discussion of the results and what they might mean, a note on limitations is perhaps due. Given the richness of the data and unlimited time and other resources, this paper could have been designed as a dynamic analysis both with respect to field maturation and the unfolding of respondents' life time and experiences as recorded in the interview (see the previous discussion of the age effect). With regard to the latter, the principal difficulty of, I think, most studies in this field lays in the fact that the data used to study the process of data collection are simultaneously also the data produced by this process; on a conceptual level, variables can be both endogenous and exogenous with respect to data quality. In other words, what I categorized above as 'respondents' characteristics, are also the product of the interviewing process and may have an 'interviewer component.' Therefore, I restricted the 'respondent part' of the design to measures for which I can reasonably assume that the 'interviewer component' tends to be absent.[60] The starting point of the present paper, as it were, is the impression that this might be less true for the dates in the event sequence. Therefore, I relied mainly on cross-sectional and admittedly rough measures characterizing respondents. In terms of durations, I used only the last transition as a proxy for memory retrieval processes.

A cautious operationalization of 'respondents characteristics' notwithstanding, a disadvantage of the design is that, in both equations, these effects have multiple 'layers of meaning.' With respect

---

[60]For example, respondents were sampled on the basis of their birth year in the context of large-scale national study with a pre-screening process. The first question in the life course interview asked about the birth year to filter out anybody not belonging to the target group, so that it is unlikely that the variable age is far off the 'true value.' Similarly, education and gender are things people are asked about all the time and in different contexts.

to gender, for example, a hypothesis could be constructed that there is something about gender which affects data quality, e.g., 'difference of salience of events' between males and females. A second layer of meaning inherent in a gender effect may be conceived of as some property of gender which leads interviewers to interact differently with male and female respondents, thereby producing differences in data quality. This latter hypothesis may be tested by an introduction of interaction terms into the model. To my knowledge, however, we do not have a theory of what such interaction effects, if any, may mean and how they come about (see Hoag & Allerbeck 1984 for a discussion of the literature).

To complicate matters further, in the model for invisible errors, the gender effect, to remain with this example, can be interpreted on yet another level of meaning. The corresponding hypothesis would be that some aspect of the social construction of gender, such as an (unconscious) disregard for female labor force participation, may lead editors to be more likely to add job spells to interviews with male respondents. Based on my experiences as supervisor of the editing group I can make the claim that this was not the case--yet in the framework of the causal analysis presented below, the interpretation of given effects is open to alternative explanations corresponding to levels of meaning.

An analysis of 'editing effects'[61] addressing this ambiguity in the interpretation of the effects on data quality found in the analysis presented here would be desirable but poses great practical and conceptual difficulties. First of all, each case was reviewed twice, and, as a rule, the second reviewer would be a different person than the first. Data entry of the changes was made only after the second review was completed, so that I have no way of telling who added which job spell other than going back to the printouts and try to identify editors by their handwriting. Second, within the small and fluctuating group of editors a division of labor was inevitable and organized around working hours and experience with editing; those who worked in the evening hours would specialize on call-backs which, in turn, were largely motivated by manifest or suspected inconsistencies in the event history

[61]Like interviewing, data editing is a social process and thus open to similar processes than those analyzed here. One possible editing effect explained above was the positive effect of tape recording on invisible errors: given a tape, editors were more likely to add job spells since tapes were the major source of information for this kind of editing. In many cases without tapes, only call-backs could provide sufficient information. Analyzing editing effects would follow the same design as the analysis of interviewer effects. Dummy variables indicating individual editors might show differences between their respective "editing style" and thoroughness, for example. Conceivably, some properties of respondents might influence the propensity of editors to detect and correct errors. The occurrence of editing errors can be modeled only for invisible errors since all visible errors were eliminated in the editing process. For visible errors, only the way in which errors were corrected (i.e. according to the rules specified in the editing handbook or not) could be analyzed.

(the discount rate for long-distance calls starts at 6 pm in Germany). The most problematic cases would usually end up on my desk. Hence, it is not possible to separate these selection effects from 'person' effects.

The ambiguity of what effects may mean is, however, not a problem in regard to variables describing aspects of interviewing from an organizational point of view. Thus, respondents could hardly influence interviewer experience, productivity, and shift size. Furthermore, even though editors could look up the interview date, editing and interviewing was not temporally synchronized, and editors could hardly know an interviewer's weekly output or the shift size.

The data I am using were not produced with a focus on methodological analysis. Rather, we incorporated related design features were it was possible to do so at low cost. A lot of measures are an offshoot of work intended to monitor and support field work and data editing. The check-program for the event sequence might have been designed differently if the present paper had been the only or even the main consideration.[62] Measurement error is associated with three variables pertaining to interviewing. Previously completed interviews, weekly productivity, and number of interviewers per shift may all be affected by the fact that the interview date was kept in the data base only for the day interview was finished; hence, interviews which had to be conducted in more than one session would carry only the date of the last session. Since the generation of these three variables is based solely on the interview date, (a) some parts of an interview, including the event sequence, may have been recorded before, (b) an interviewer who only started but did not complete interviews during a shift would not show up in the data. It is not possible to systematically control the extent of measurement error since field protocols are not in machine readable form; regular inspection during the field time showed that almost all interrupted interviews were completed within a week of the first session.[63]

---

[62]For example, we could have observed data quality as it evolves over the event sequence (its 'life-time' dependency), or measured the clustering of errors within events as well as the length of gaps and overlaps to obtain additional information.

[63]However, I found a few cases where the interview was stretched over several months! Deleting these and other potential outliers and influential cases from the analysis did not affect the conclusions drawn from the results. See above in the section on outlier analysis.

VI. CONCLUSION

The results presented above are, in many respects, not generalizable to survey research at large. A comparison of data quality among longitudinal surveys with different designs is needed to evaluate the data in perspective. The extent to which they provide a contribution to the field of survey methodology depends in part on where the trend in survey research will go. Virtually all surveys work in part with retrospective data collection but rarely as extensive as in the Life History Study considered here. Hardly any survey has similarly high interviewer workloads yet many institutions work with more or less stable interviewer staffs, and a professionalization of interviewing has been discussed in the social sciences for a long time.

The first question I will address is whether the data analyzed here are affected by an 'interviewer bias.' My results speak to differences in data quality and therefore to the comparability of cases across interviewer samples. I have shown that interviewers may influence the number of job spells coded in an interview (see Appendix F). I find 5 significant coefficients for individual interviewers, with effect sizes between 17 and 30%. These interviewers recorded less job spells than others, controlling for other variables. In addition, the more interviewers work, the more job spells we find in the unedited interviews. These effects were eliminated during data edition. The same analysis performed after data editing shows no effects of interviewing or individual interviewers on the number of job spells.[64]

Individual interviewers certainly do influence the number of errors counted by my measures. Partly, data editing can remove these influences, but may introduce editing effects such as the fact that more job spells were added for interviews with tape recording. Also, data editing is costly and error prone. Where call-backs were impossible, where no tape recording could provide additional information, missing data on the event sequences had to be imputed.

Interviewer effects on interview time can be both beneficial and harmful with respect to data quality. First, interviewers may become more efficient in leading respondents through the questionnaire, avoiding confusion and supporting the retrieval process. Also, where life courses are complex with respect to the instrument, and depending on the volume of the information transmitted, additional time enhances data quality. Second, interviewers may push respondents through the event

---

[64]As explained above, a comparison of visible errors before and after data editing is not useful since all visible errors were eliminated. One could compute the duration of events before and after editing in order to make that comparison. Still, for events with missing data on the event sequence (see table I) no duration can be calculated.

sequence, fail to probe for answers, and skip questions, thereby saving time but risk errors. The regression of logged interview time reported in Appendix D shows that almost all of the 19 most active interviewers take significantly less time than the five interviewers who completed less than 30 interviews. Experience also has a negative effect on interview time, net of other variables. Furthermore, the results from a logistic regression on tape recording reported in Appendix E show that while all but five of the 19 interviewers produce more interviews with tape recording than the comparison group, with growing experience and productivity they do so less and less. My results thus point to a complex web of effects on data quality where experience, routinization, field control with tape recordings, and, last but not least, individuals, have direct and/or indirect effects on data quality which partially operate through their effects on interview time, which, in turn, may be conditional on volume and complexity of the interviews. My results should not be taken, however, to mean that interviewers with work loads below 30 cases do not have any impact on data quality--rather, showing such effects with statistic techniques is more difficult or impossible as N's get smaller.

The Life History Study in its 10 year-long development confronted the many-fold problems of retrospective data collection with a switch from face-to-face interviewing to the computer-assisted telephone mode; the high standards of data quality set by the development of event history analysis were increasingly met by the development of techniques for data editing and call-backs. In response to problems with response rates and data quality, interviewer staffs became smaller and smaller, and much was invested in training and supervision. Within this context, I think my results show that (1) the mechanisms of field control in terms of interviewer selection and supervision worked imperfectly, both with respect to field trends and to individual interviewers, and that (2) interviewer experience and productivity are mixed blessings.

Certainly, the field staff became more cost efficient with respect to the ratio of collected information and interview time. In part, however, this increase in efficiency was accompanied by an increase in visible errors. The fact that it did not result also in a decline of meaningfulness may confirm the impression we got during the field period, namely that, generally speaking, interviewers were highly committed to the survey and did their best to 'get the story across.' I think that I do not stretch my results too far by saying that at least some interviewers traded off 'visible data quality' for 'invisible data quality' or meaningfulness of data. In future surveys, the importance of collecting the boring and tedious 'dates' of events should be communicated even better. Perhaps a greater investment in feed-back and de-briefing could have helped to avoid or attenuate the increase of visible errors. The greatest emphasis on feed-back was naturally planned for the first phase of field work. An emphasis on the importance of the 'dry' facts during interviewer training and continued field control and intervention might thus be advisable for the future.[65]

Had we been able to detect this tendency during the field period, specific interventions might have helped to reverse the growth of visible errors. In other words, the potential of the CATI method for timely systematic field control of the directly entered data was not fully realized in this survey. Given time, know-how, and resources, we could have developed a computer-assisted routine for a direct control of data quality like the check program we designed later for the support of the manual editing.

The interaction effect of interview time and volume driven by the inclusion of 'difficult' respondents in the model for invisible errors may speak to the effect that the Life History Study has pushed data collection to the limits of what survey research can accomplish with presently available methods and technology. Furthermore, the effect of life course complexity especially with respect to invisible errors points out two problems future research should worry about. First, are complex life stories harder to collect, and may lead to incomplete data? My paper provides some support for this. Second, the measure indicating complexity was designed to enumerate transitions between life domains. Given the different details and measurement concepts we used for each domain, one of the earliest developments in the Life History Study was the decision to break up the chronological sequence of events into these domains.

This method worked well with face-to-face interviews (Brückner 1990, 1994) but my results may indicate that, at least in these cohorts, the telephone mode may result in data quality problems when complex life stories are encountered. Intuitively, complex transitions are easier to master when going back and forth between the different domains is possible. In the CATI system, technical problems precluded going back so that everything had to be right at the first try. This conclusion is tentative, of course, and comparisons between face-to-face and telephone interviews are needed to examine the possible interaction between the survey mode and life course complexity.

In addition, the negative impact of interruptions in the interview on data quality may be a true survey mode effect; face-to-face interviews conducted in two other parts of the Life History Study were almost never interrupted. On the other hand, this paper was only possible because of the CATI

[65] In June 89, we met the interviewer staff for additional briefing. Tapes of interviews with inconsistencies were used to draw interviewers' attention to the operationalization of job changes and special emphasis was laid on incomplete job histories. The steep decline in the number of job spells added during data editing for interviews conducted in July 89 (see figure 5) may or may not be a consequence of this intervention.

method, which allowed the analysis of the raw data and a comparison of edited and unedited data. My mother used to say that the field is "a black box for the researcher." Direct data entry and processing with CATI, however, makes detailed and intensive studies of field processes possible without weighing too heavily on the survey budget.

I have said in the introduction that the developments in the social sciences continually push the need for advances in data collection methods; I might add here that systematic research about data quality, however painful and difficult it may be, is needed in times when the confidence in survey research declines not only in the general public (see Groves 1989) but also within the profession itself.

In regard to respondent effects, my results are to be interpreted with caution because effects may have multiple meanings. Perhaps these concerns are better addressed within an experimental research design. Suffice it to say that the strong effects of respondents' education replicate results of previous research and should be taken as an incentive to work on methods of data collection which are better suited to less well educated people. The inadequacy of survey questions in this respect is standard 'shop talk' among interviewers who readily admit that they change the wording of questions because 'respondents just don't get it.' This is, of course, in part a justification of negatively sanctioned behavior--but it still rings true when thinking back to the many situations overheard in the tapes where respondents completely missed the stimulus of a question even after several repetitions.

I think that most of survey research lives with the tacit or open acceptance of the adaptations interviewers introduce into questionnaires when encountering such problems (Bradburn 1988); it is an uneasy life, however, because these 'adaptations' can go right and wrong, depending on interviewers' skill and understanding of the measurement concept. More often than not, things may go wrong because interviewers have normally no guidelines how to react to such problems since, of course, no researcher wants to train interviewers to deviate from the questionnaire. Hence, interviewers may be undertrained especially where more training is needed to deal with interactional trouble in the interview.

The life history approach will continue to heighten our understanding of the dynamics of social change and thus merits every possible effort to understand and improve the collection of life history data. There is some support for the notion that data on labor force participation is easier to collect than data on, say, migration, schooling, vocational training, home-making and unemployment spells. And how might retrospective data collection work in a context were people are less obsessed with documentation and certification of life events than in Germany? My paper provides support not only for the importance of the nature of survey organization for the evaluation of survey data quality. In addition, my results carry some support for the notion that a conceptualization of data quality which takes into account both appearance and meaningfulness of data may provide valuable insights into the dynamics of interviewing.

Appendix A: Baseline Model for Visible Errors. Comparison of Negative Binomial and L-OLS Estimates.

Variable	Negative Binomial coefficients	s.e.	L-OLS coefficients	s.e.
Constant	-3.0462*	1.3876	-1.9595	1.3495
Interview				
Log of time	.6577*	.3278	.4468	.3250
Volume	.1894	.3430	.0038	.3358
Log time*volume	-.1178	.0795	-.0634	.0796
Log # of events	1.4844**	.0604	1.3533**	.0534
# of job spells	-.0354**	.0061	-.0337**	.0070
Interviewing				
Tape recording	-.0328	.0206	-.0336*	.0199
Experience	.0011**	.0003	.0010**	.0003
Productivity	.0124	.0072	.0178*	.0069
# interviewers in shift	-.0026	.0038	-.0039	.0037
Two sessions	.0668**	.0192	.0323	.2050
Three + sessions	.1066*	.0509	.0850	.0571
Respondent				
Female	.0068	.0177	.0064	.0174
Age	.0121**	.0036	.0120**	.0034
Time since transition	.0081	.0127	.0021	.0119
Frequency	-.0058	.0055	-.0089	.0054
Complexity	-.0039	.0081	-.0001	.0084
Hauptschule	-.0958*	.0452	-.0910*	.0444
Middle School	-.2418**	.0469	-.2286**	.0456
Fachhochschulreife	-.2929**	.0531	-.2867**	.0525
Abitur	-.3366**	.0477	-.3208**	.0470
Living with partner	-.0036	.0208	.0017	.0203
# of small children	-.0214	.0137	-.0108	.0116
Individual Interviewers				
1	-.0079	.0715	.0106	.0660
2	-.1975**	.0578	-.2404**	.0589
3	.1349*	.0618	.1483**	.0569
4	-.3931**	.0541	-.4108**	.0544
5	.3396**	.0553	.3172**	.0454
6	-.0252	.0869	.0294	.0661
7	-.1287*	.0583	-.1157*	.0530
8	-.0158	.0578	-.0053	.0542
9	-.1247*	.0544	-.1244*	.0533
10	.2116**	.0576	.2009**	.0574
11	.0002	.0726	.0178	.0599
12	-.0116	.0568	.0121	.0531
13	.0016	.0691	.0354	.0588
14	.2750**	.0686	.2652**	.0695
15	.1221	.0681	.1302	.0702
16	.0488	.0637	.0835	.0561
17	.0396	.0540	.0463	.0526
18	.0306	.0771	.0751	.0624
19	.0407	.0596	.0413	.0621
Dispersion parameter α	.0407**	.0032	--	
Adjusted R^2	--		.585	
Likelihood ratio	1912.4 with 41 df		--	

*$p \leq .05$ **$p \leq .01$

Appendix B: Poisson Regression of Invisible Errors including visible Errors as Regressor. Comparison with L-OLS Estimates.

Variable	Poisson coefficients	s.e.	L-OLS coefficients	s.e.
Constant	2.7304*	8.5383	.9670	1.1205
Interview				
Log of visible errors	.0775	.1418	.0380	.0188
Log of time	-5.1164*	2.0766	-.7272*	.2698
Volume	1.5158	2.0205	.2591	.2786
Log time*volume	.9696*	.4895	.1380*	.0661
Log # of events	-2.8160**	.3279	-.4989**	.0511
# of job spells	-.3335**	.0415	-.0485**	.0058
Interviewing				
Tape recording	.2558*	.1252	.0173	.0165
Experience	-.0028	.0015	-.0005*	.0002
Productivity	-.0805*	.0365	-.0102	.0058
# interviewers in shift	.0119	.0213	.0039	.0030
Two sessions	.1198	.1196	.0249	.0170
Three + sessions	.5891*	.2691	.1284**	.0437
Respondent				
Female	-.8145**	.1077	-.1166**	.0145
Age	.1070**	.0202	.0134**	.0028
Time since transition	-.3551**	.0606	-.0623**	.0099
Frequency	.0514	.0369	.0038	.0045
Complexity	.1268**	.0496	.0240**	.0069
Hauptschule	-.2980	.2029	-.0551	.0369
Middle School	-.2895	.2176	-.0601	.0381
Fachhochschulreife	-.7261**	.2583	-.1071*	.0487
Abitur	-.7936**	.2280	-.1355**	.0395
Living with partner	-.1881	.1210	-.0661	.0429
# of small children	-.0399	.0671	-.0103	.0097
Individual Interviewers				
1	.7791	.4538	.0800	.0547
2	-.2572	.4795	-.0271	.0491
3	.2670	.4378	.0093	.0473
4	-.2495	.4450	-.0283	.0458
5	1.2269**	.4034	.1663**	.0456
6	-.2866	.5788	-.0106	.0549
7	.5412	.4136	.0721	.0440
8	.5335	.4114	.0858	.0449
9	.4726	.4117	.0661	.0443
10	.9130*	.4105	.1320**	.0478
11	.9948*	.4213	.1055*	.0497
12	.2237	.4114	.0375	.0441
13	-.6525	.5757	-.0596	.0488
14	.9291*	.4481	.1346*	.0579
15	.6457	.4683	.0545	.0583
16	.3223	.4340	.0022	.0466
17	.3113	.4048	.0436	.0437
18	.7383	.4300	.0682	.0518
19	.8187	.4495	.0998	.0516
R^2	--		.23	
Likelihood ratio	678.8 with 42 df		--	

*p ≤ .05 **p ≤ .01

Appendix C: Poisson Regression of Invisible Errors Including Visible Errors as Regressor. Models with Full Sample and Outliers Removed.

Variable	Full Sample (N=1987) coefficients	s.e.	Outlier deleted (N=1854) coefficients	s.e.
Constant	2.7304*	8.5383	-38.6667	17.5326
Interview				
Log of visible errors	.0775	.1418	.0987	.2215
Log of time	-5.1164*	2.0766	4.3724	4.2087
Log # of records	1.5158	2.0205	11.8340*	4.3437
Log time*log records	.9696*	.4895	-1.4251	1.0318
Log # of events	-2.8160**	.3279	-3.0105**	.5143
# of job spells	-.3335**	.0415	-.5499**	.0758
Interviewing				
Tape recording	.2558*	.1252	.1901	.1882
# prev. completed cases	-.0028	.0015	-.0042*	.0021
weekly output	-.0805*	.0365	-.0658	.0547
# interviewers in shift	.0119	.0213	.0165	.0322
Two sessions	.1198	.1196	.0026	.1866
Three + sessions	.5891*	.2691	.3408	.5405
Respondent				
Female	-.8145**	.1077	-.9326**	.1609
Age	.1070**	.0202	.1167**	.0300
Time since transition	-.3551**	.0606	-.4619**	.0886
Frequency	.0514	.0369	.1554*	.0612
Complexity	.1268**	.0496	.2130*	.0886
Hauptschule	-.2980	.2029	.0437	.3670
Middle School	-.2895	.2176	-.1564	.3884
Fachhochschulreife	-.7261**	.2583	-1.0004*	.4890
Abitur	-.7936**	.2280	-.8819*	.4187
Living with partner	-.1881	.1210	-.1819	.1920
# of small children	-.0399	.0671	-.0431	.1019
Individual Interviewers				.8304
1	.7791	.4538	1.4141	.8249
2	-.2572	.4795	.6678	.8073
3	.2670	.4378	.7516	.8199
4	-.2495	.4450	.2913	.7681
5	1.2269**	.4034	1.8438*	.9311
6	-.2866	.5788	.3496	.7714
7	.5412	.4136	1.1972	.7633
8	.5335	.4114	1.3986	.7680
9	.4726	.4117	1.3012	.7655
10	.9130*	.4105	2.0496**	.7971
11	.9948*	.4213	1.4707	.8108
12	.2237	.4114	.3053	1.0093
13	-.6525	.5757	-.3353	.7969
14	.9291*	.4481	1.7908*	1.0147
15	.6457	.4683	.2360	.9229
16	.3223	.4340	.1095	.7603
17	.3113	.4048	.9875	.7965
18	.7383	.4300	1.2010	1.0180
19	.8187	.4495	.3131	
Loglikelihood Ratio	678.4 with 41 df		333.6 with 41 df	

*p ≤ .05   **p ≤ .01

Appendix D: OLS Regression of selected independent Variables on Logged Interview Time.

Variable	coefficients	s.e.
Constant	1.6642**	.0937
Interview		
Log # of visible errors	.0561**	.0113
Volume	.5612**	.0292
Log # of events	.0394	.0311
# of job spells	.0218**	.0034
Interviewing		
Tape recording	.0228*	.0010
Experience	-.0013**	.0001
Productivity	.0007	.0035
# interviewers in shift	.0078**	.0018
Respondent		
Female	.0080	.0086
Age	.0001	.0017
Time since transition	.0145*	.0059
Trajectories:		
Frequency	-.0050	.0027
Complexity	.0045	.0041
Schooling: degree	.0028	.0040
Living with partner	-.0049	.0102
# of small children	-.0073	.0063
Individual Interviewers		
1	-.3519**	.0323
2	-.2874**	.0290
3	-.3694**	.0275
4	-.3125**	.0268
5	-.0197	.0276
6	-.2790**	.0327
7	-.0718**	.0266
8	-.1905**	.0269
9	-.0788**	.0268
10	-.2706**	.0284
11	-.0304	.0301
12	-.1161**	.0266
13	-.1589**	.0294
14	-.1375**	.0350
15	-.2452**	.0349
16	-.0210	.0283
17	-.2504**	.0259
18	-.0556	.0315
19	.3275**	.0304
F-Value	117.57**	
Adjusted R^2	.67	

*$p \leq .05$ **$p \leq .01$

Appendix E: Logistic Regression of Selected Variables on Tape Recording.

Variables	coefficients	s.e.
Constant	.9077	.8064
Interviewing		
Experience	-.0088**	.0019
Productivity	-.1097*	.0485
# interviewers in shift	-.0117	.0256
Respondent		
Female	-.2190	.1195
Age	.0301	.0221
Trajectories:		
Frequency	-.0195	.0363
Complexity	.0237	.0548
Schooling: degree	-.0819	.0498
Living with partner	.0194	.1345
# of small children	.1389	.0866
Individual Interviewers		
1	-1.2100**	.3784
2	.9439*	.4008
3	-.6679*	.3263
4	1.2005**	.3572
5	1.8963**	.3848
6	.6703	.4473
7	1.7440**	.3697
8	1.8839**	.3879
9	1.0137**	.3476
10	.0192	.3478
11	.4997	.3875
12	.6798*	.3427
13	.1916	.3663
14	-.2745	.4121
15	-.9651*	.4035
16	.6185	.3701
17	-.2802	.3164
18	1.3271**	.5033
19	1.3581**	.4731
chi-square for covariates	228.3 with 29 df	

*$p \leq .05$ **$p \leq .01$

Appendix F: Poisson Regression of Selected Variables on Number of Job Spells before and after Data Editing.

	Number of jobs before edition		Number of jobs after edition	
Variables	coefficients	s.e.	coefficients	s.e.
Constant	.4209*	.2122	.4000	.2094
Interviewing				
tape recording	.0609*	.0337	.0533	.0331
# prev. completed cases	.0014**	.0004	.0007	.0004
weekly output	.0040	.0110	-.0079	.0106
# interviewers in shift	.0047	.0061	.0023	.0060
Respondent				
Female	.0068	.0504	-.0446	.0502
Age	.0338**	.0054	.0409**	.0053
Schooling: degree	-.0589**	.0206	-.0830**	.0202
Schooling: duration	-.0019	.0011	-.0024*	.0011
Living with partner	.0252	.0456	.0580	.0444
# of small children	-.0097	.0280	-.0005	.0268
Female*living w. partner	-.1715**	.0648	-.1887**	.0640
Female*children	.0010	.0414	-.0031	.0405
Individual Interviewers				
1	-.0093	.1062	.0700	.1055
2	-.1183	.0968	-.0455	.0971
3	-.0768	.0912	-.0102	.0919
4	-.0392	.0854	.0148	.0863
5	-.1841*	.0906	-.0249	.0904
6	.1103	.1017	.1122	.1033
7	-.1322	.0879	-.0256	.0884
8	-.1826*	.0895	-.0067	.0891
9	-.0917	.0879	.0419	.0880
10	-.2058*	.0957	-.0252	.0937
11	-.1091	.1008	.0152	.0996
12	-.1085	.0879	-.0067	.0881
13	-.1141	.0977	-.0779	.0985
14	-.3339**	.1273	-.2218	.1251
15	-.0538	.1161	.0539	.1138
16	-.1692	.0950	-.0923	.0951
17	-.2086	.0867	-.0769	.0867
18	-.3516**	.1134	-.1937	.1100
19	-.0575	.0999	.0515	.0991
Likelihood ratio	323.6 with 31 df		203.2 with 31 df	

*$p \leq .05$ **$p \leq .01$

REFERENCES

Allmendinger, Jutta. 1989. Career Mobility Dynamics. A Comparative Analysis of the United States, Norway, and West Germany. Berlin: Edition Sigma, Reihe Studien und Berichte des Max-Planck-Instituts für Bildungsforschung Nr. 49.

Argyris, Chris. 1980. Inner Contradictions of Rigorous Research. New York/London: Academic Press.

Argyris, Chris. 1968. Some Unintended Consequences of Rigorous Research. Psychological Bulletin, 70, pp. 185-197.

Atteslander, Peter. 1984. Methoden der empirischen Sozialforschung. Berlin/New York: Walter de Gruyter.

Bailey, L., T. Moore, and B. A. Bailar. 1978. An Interviewer Variance Study for the Eight Impact Cities of the National Crime Survey Cities Sample. Journal of the American Statistical Association, 73, pp. 16-23.

Bateson, Nicholas. 1984. Data Construction in Social Surveys. London: George Allen & Unwin.

Berk, Marc L., and Amy B. Bernstein. 1988. Interviewer Characteristics and Performance on a Complex Health Survey. Social Science Research, 17, pp. 239-251.

Bernard, Russel H., Peter Killworth, David Kronenfeld, and Lee Sailer. 1984. The Problem of Informant Accuracy: The Validity of Retrospective Data. Annual Review of Anthropology, 13, pp. 495-517.

Bertaux, Daniel (ed.) 1981. Biography and Society. The Life History Approach in the Social Sciences. Beverly Hills: Sage

Bertaux, Daniel, and Martin Kohli. 1984. The Life Story Approach: A Continental View. Annual Review of Sociology 10, pp. 215-237.

Billiet, Jacques, and Geert Loosveldt. 1988. Improvement of the Quality of Responses to Factual Survey Questions by Interviewer Training. Public Opinion Quarterly, 52, pp. 190-211.

Blum, Zahava, Nancy Karweit, and Aage B. Sørensen. 1968. A Method for the Collection and Analysis of Retrospective Life Histories. Baltimore: The John Hopkins University.

Bogdan, Robert, and Magret Ksander. 1980. Policy Data as a Social Process: A Qualitative Approach to Quantitative Data. Human Organization, 39, pp. 302-309.

Bollen, Kenneth A., and Robert W. Jackman. 1990. Regression Diagnostics: An Expository Treatment of Outliers and Influential Cases. In J. Fox and J. S. Long (eds.): Modern Methods of Data Analysis. Newbury Park, CA: Sage.

Bradburn, Norman M. 1988. Response Effects. In Peter H. Rossi, James D. Wright, and Andy B. Anderson (eds.): Handbook of Survey Research. New York: Academic Press, pp. 289-328.

Brenner, Michael. 1978. Interviewing: the Social Phenomenology of a Research Instrument. In Michael Brenner, Peter Marsh, and Marylin Brenner (eds.): The Social Contexts of Method. London: Croom Helm, pp. 122-139.

Brenner, Michael. 1982. Response Effects of "Role-restricted" Characteristics of the Interviewer. In Wil Dijkstra and Johannes van der Zouwen (eds.): Response Behavior in the Survey Interview. New York/London: Academic Press, pp. 131-165.

Brückner, Erika. 1994. Erhebung ereignisorientierter Lebensverläufe als retrospektive Längsschnittkonstruktion. In Richard Hauser, Notburga Ott, und Gert Wagner (Hrg.): Mikroanalytische Grundlagen der Gesellschaftspolitik Vol. 2. Berlin: Akademischer Verlag, pp. 38-69

Erika Brückner. 1993. Lebensverläufe und gesellschaftlicher Wandel - Konzeption, Design und Methodik der Erhebung von Lebensverläufen der Geburtsjahrgänge 1919-1921. Teile I-V. Berlin: Max-Planck-Institut für Bildungsforschung, Materialien aus der Bildungsforschung Nr. 44.

Brückner, Erika. 1990. Die retrospektive Erhebung von Lebensverläufen. In Mayer, K. U. (ed.): Lebensverläufe und sozialer Wandel. Kölner Zeitschrift für Soziologie und Sozialpsychologie, Sonderheft 31, pp. 374-403.

Brückner, Erika. 1989. Methodenreport: Feldbericht Hauptstudie. In Karl Ulrich Mayer und Erika Brückner (eds.): Lebensverläufe und Wohlfahrtsentwicklung - Konzeption, Design und Methodik der Erhebung von Lebensverläufen der Geburtsjahrgänge 1929-1931, 1939-1941, 1949-1951 Teil I. Berlin: Max-Planck-Institut für Bildungsforschung, Materialien aus der Bildungsforschung Nr. 35, pp. 123-172.

Brückner, Hannah, and Karl Ulrich Mayer (eds.). 1995. Berufszugang in der Beschäftigungskrise - Konzeption, Design und Methodik der Erhebung von Lebensverläufen der Geburtsjahrgänge 1954-1955 und 1959-1961. Teil I-III. Berlin: Max-Planck-Institut für Bildungsforschung, Materialien aus der Bildungsforschung Nr. 49.

Brückner, Hannah (1995): Anhang: Computerunterstützte Konsistenzprüfung. In Hannah Brückner and Karl Ulrich Mayer (eds.). 1995. Berufszugang in der Beschäftigungskrise - Konzeption, Design und Methodik der Erhebung von Lebensverläufen der Geburtsjahrgänge 1954-1955 und 1959-1961. Berlin: Max-Planck-Institut für Bildungsforschung, Materialien aus der Bildungsforschung Nr. 49, Teil II, pp. 213ff.

Bryant, Joseph M. 1994. Evidence and Explanation in History and Sociology: Critical Reflections on Goldthorpe's Critique of Historical Sociology. British Journal of Sociology, 45 (1), pp. 3-19.

Cameron, A. Colin, and Pravin K. Trivedi. 1986. Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests. Journal of Applied Econometrics 1, pp. 29-53.

Cannell, Charles F. und Manfred Küchler. 1984. Antwortverzerrung im Interview - Wie läßt sich die Güte der Daten verbessern? Zuma Nachrichten, Nr. 15, pp. 3-17.

Cannel, Charles F., Peter V. Miller, and Lois Oksenberg. 1981. Research on Interviewing Techniques. Sociological Methodology, 11, pp. 389-437.

Chassein, Brigitte und Hans-Jürgen Hippler. 1987. Reliabilität und Validität retrospektiver Daten. Befunde aus der kognitiven Psychologie (Sektion Biographieforschung i.d. DGS), in Jürgen Friedrichs (ed.): Technik und Sozialer Wandel, Opladen, pp. 453-456.

Clarren, S., and A. Schwartz. 1976. Measuring a program's impact: A cautionary note. In W. Skogan (ed.): Sample Surveys of the Victims of Crime. Cambridge, Mass.: Ballinger.

Converse, Jean M., and Stanley Presser. 1986. Handcrafting the Standardized Questionnaire. Newbury Park, CA: Sage.

Converse, Jean M. and Howard Schuman. 1974. Conversations at Random. New York: John Wiley & Sons.

Denzin, Norman K. (ed.). 1978. Sociological Methods: A Sourcebook. Second Edition. New York: McGraw-Hill.

Deutscher, Irwin. 1978. Asking Questions Cross-Culturally: Some Problems of Linguistic Comparability. In Norman K. Denzin (ed.): 1978. Sociological Methods: A Sourcebook. Second Edition. New York: McGraw-Hill, p. 182-204.

Deutscher, Irwin. 1972. Public and Private Opinions: Social Situations and Multiple Realities. In S. Z. Nagi and R. G. (eds.): The Social Contexts of Research. London: Wiley.

Dex, Shirley. 1971. The Reliability of Recall Data: A Literature Review. Working Papers of the ESRC Research Centre on Micro-Social Change, No. 11. University of Essex: Economic and Social Research Council.

Dijkstra, Wil. 1987. Interviewing Style and Respondent Behavior. An Experimental Study of the Survey Interview. Sociological Methods and Research, 16, pp. 309-334.

Dijkstra, Wil, and Johannes van der Zouwen (eds.): 1982. Response Behavior in the Survey Interview. New York/London: Academic Press.

Douglas, Jack D. 1985. Creative Interviewing. Beverly Hills, C.A.: Sage.

Duncan, Greg J., and John N. Morgan. 1985. The Panel Study of Income Dynamics. In Glen H. Elder, (ed): Life Course Dynamics. Trajectories and Transitions, 1968-1980. Ithaca and London, Cornell University Press.

Elder, Glen H. 1985. Perspectives on the Life Course. In Elder, Glen H. (ed.): Life Course Dynamics. Trajectories and Transitions, 1968-1980. Ithaca and London, Cornell University Press, pp. 21-49.

Esser, Hartmut. 1984. Determinanten des Interviewer- und Befragtenverhaltens: Probleme der theoretischen Erklärung und empirischen Untersuchung von Interviewereffekten. In Karl Ulrich Mayer and Peter Schmidt (eds.): Allgemeine Bevölkerungsumfrage der Sozialwissenschaften. Frankfurt (Main)/New York: Campus, pp. 26-71.

Featherman, D. L. 1979. Retrospective Longitudinal Research: Methodological Considerations. Journal of Economics and Business 32, pp. 152-169.

Freedman, Deborah, Arland Thornton, Donald Camburn, Duane Alwin, and Linda Young-DeMarco. 1988. The Life History Calendar: A Technique for Collecting Retrospective Data. Sociological Methodology 18, pp. 435-463.

Freeman, Howard E. 1983. Research Opportunities Related to CATI. Sociological Methods and Research, 12, pp. 143-152.

Freeman, John, and Edgar W. Butler. 1976. Some Sources of Interviewer Variance in Surveys. Public Opinion Quarterly, 40, pp. 79-91.

Greene, William H. 1990. LIMDEP. New York: Econometric Software, Inc. Revised Edition.

Furstenberg, Frank L., J. Brooks-Gunn, and S. Philip Morgan. 1987. Adolescent Mothers in Later Life. New York: Cambridge University Press.

Gibson, C. O., G. M. Shapiro, L. R. Murphy, and G. J. Stanko. 1978. Interaction of Survey Questions as it Relates to Interviewer-Respondent Bias. Proceedings of the American Statistical Association (Survey Research Methods Section), 1978, pp. 251-256.

Groves, Robert M., and Lou J. Magilavy. 1986. Measuring and Explaining Interviewer Effects in Centralized Telephone Surveys. Public Opinion Quarterly, 50, 251-266.

Groves, Robert M. 1990. Theories and Methods of Telephone Surveys. Annual Review of Sociology, 16, pp. 221-240.

Groves, Robert M. 1989. Survey Error and Survey Costs. New York: Wiley.

Groves, Robert M. 1987. Research on Data Quality. Public Opinion Quarterly, 51, pp. S156-S172.

Groves, Robert M., Paul B. Biemer, Lars E. Lyberg, James T. Massey, William L. Nicholls II, and Joseph Waksberg (eds.). 1987. Telephone Survey Methodology. New York: John Wiley & Sons.

Groves, Robert M., and Nancy H. Fultz. 1985. Gender Effects among Telephone Interviewers in a Survey of Economic Attitudes. Sociological Methods and Research, 14, pp. 31-52.

Groves, Robert M., and Robert L. Kahn. 1979. Surveys by Telephone. A National Comparison with Personal Interviews. New York/London: Academic Press.

Hanson, Robert H., and Eli S. Marks. 1958. Influence of the Interviewer on the Accuracy of Survey Results. Journal of the American Statistical Association, 53, pp. 635-655.

Hatchett, S., and H. Schuman. 1975. White Respondents and Race-of-Interviewer Effects. Public Opinion Quarterly, 39, pp. 523-528.

Helling, Ingeborg. 1987. Methodische Überlegungen zum Einsatz von Fallanalysen am Beispiel Arbeitserfahrung und Berufspositionenfolge. In Wolfgang Voges (ed.): Methoden der Biographie- und Lebenslaufforschung. Opladen, pp. 197-230.

Hoag, Wendy J., and Klaus R. Allerbeck. 1981. Interviewer- und Situationseffekte in Umfragen: Eine log-lineare Analyse. Zeitschrift für Soziologie, 10 (4), pp. 413-426.

Huttenlocher, Janellen, Larry Hedges, and Vincent Prohaska. 1988. Hierarchical Organization in Ordered Domains: Estimating the Dates of Events. Psychological Review, 95, pp. 471-487.

Hyman, Herbert H. and Associates. 1954. Interviewing in Social Research. Chicago: The University of Chicago Press.

King, Gary. 1989. Variance Specification in Event Count Models: From Restrictive Assumptions to a Generalized Estimator. American Journal of Political Science 33, pp. 762-784.

King, Gary. 1988. Statistical Models for Political Science Event Counts: Bias in Conventional Procedures and Evidence for the Exponential Poisson Regression Model. American Journal of Political Science 32, pp. 838-863.

Krupp, Hans-Jürgen, and Ute Hanefeld (eds.). 1987. Lebenslagen im Wandel: Analysen 1987. Frankfurt/Main: Campus.

Lipset, S. Martin. 1992. Polls Don't Lie. People Do. New York Times, September 10th, p. A17.

Loftus, Elizabeth F. und Wesley Marburger. 1982. Memory and its Distortions. In Alan G. Kraut (ed.): The G. Stanley Hall Lecture Series. Volume 2. American Psychological Association, Washington D.C.

Loftus, Elizabeth F. und Wesley Marburger. 1983. Since the Eruption of Mt. St. Helens, Has Anyone Beaten You up? Improving the Accuracy of Retrospective Reports with Landmark Events. Memory & Cognition, 11, pp. 114-120.

Marquis, Kent H. 1969. A Study of Interviewer-Respondent Interaction in the Urban Employment Survey. The University of Michigan, Survey Research Center.

Martin, Elizabeth. 1988. Surveys as Social Indicators: Problems in Monitoring Trends. In Peter H. Rossi, James D. Wright, and Andy B. Anderson (eds.): Handbook of Survey Research. New York: Academic Press, pp. 677-743.

Mayer, Karl Ulrich und Johannes Huinink. 1990. Age, Period and Cohort in the Study of the Life Course: A Comparison of Classical A-P-C Analysis with Event History Analysis or Farewell to Lexis?, in David Magnusson and Lars R. Bergman (eds.): Methodological Issues in Longitudinal Research. Cambridge: Cambridge University Press, pp. 211-232.

Mayer, Karl U., und Erika Brückner (eds.). 1989. Lebensverläufe und Wohlfahrtentwicklung. Konzeption, Design und Methodik der Erhebung von Lebensverläufen der Geburtsjahrgänge 1929-1931, 1939-1941, 1949-1951. Teile I-III. Berlin: Max-Planck-Institut für Bildungsforschung, Materialien aus der Bildungsforschung Nr. 35.

Mayer, Karl U., and Peter Schmidt. 1984. Allgemeine Bevölkerungsumfrage der Sozialwissenschaften. Frankfurt (Main)/New York: Campus.

McCullagh, Peter, and John A. Neder. 1983. Generalized Linear Models. London, New York: Chapman and Hall.

Merton, Robert K., Majorie Fiske, and Patricia L. Kendall. 1956. The Focused Interview. A Manual of Problems and Procedures. Glencoe, Ill.: The Free Press.

Niethammer, Lutz. 1989. Annäherung an den Wandel. Auf der Suche nach der volkseigenen Erfahrung in der Industrieprovinz der DDR. In Alt Lüdtke (ed.): Alltagsgeschichte. Zur Rekonstruktion historischer Erfahrungen und Lebensweisen. Frankfurt (Main), pp. 289ff.

Nuthmann, Reinhard, and Hannah Brückner (1995): Editionshandbuch zur Lebensverlaufsstudie III - Kohorten 1954-56 und 1959-61. In Hannah Brückner and Karl Ulrich Mayer (eds.). 1995. Berufszugang in der Beschäftigungskrise - Konzeption, Design und Methodik der Erhebung von Lebensverläufen der Geburtsjahrgänge 1954-1955 und 1959-1961. Berlin: Max-Planck-Institut für Bildungsforschung, Materialien aus der Bildungsforschung Nr. 49, Teil II, pp. 213ff.

Oppenheimer, Valerie K. 1974. The Life-cycle Squeeze: The Interaction of Men's Occupational and Family Cycles. Demography 11, pp. 237-245.

Pennef, Jean. 1988. The Observers Observed: French Survey Researchers at Work. Social Problems, 35, pp. 520-535.

Prüfer, Peter, and Magrit Rexroth. 1985. Zur Anwendung der Interaction-Coding-Technik. Zuma-Nachrichten, 17, pp. 2-47.

Rice, S. A. 1929. Contagious Bias in the Interview. American Journal of Sociology, 35, pp. 420-423.

Riesman, David. 1958. Some Observations on the Interviewing in the Teacher Apprehension Study. In Paul L. Lazarsfeld, and Wagner Thielsen: The Academic Mind. Glencoe, Ill.: The Free Press, pp. 266-370.

Rindfuss, Ronald R., C. Gray Swicegood, and Rachel A. Rosenfeld. 1987. Disorder in the Life Course: How Common Is It and Does It Matter? American Sociological Review, 52, pp. 785-801.

Rosenthal, Robert. 1969. Interpersonal Expectations: Effects of the Experimenter's Hypothesis. In Robert Rosenthal and Ralph L. Rosnow (eds.): Artifact in Behavioral Research. New York: Academic Press, pp. 182-277.

Roth, Julius. 1966 [1978]. Hired Hand Research. The American Sociologist, 1, pp. 190-196. Reprinted in Norman K. Denzin (ed.[1978]): Sociological Methods: A Sourcebook. Second Edition. New York: McGraw-Hill, p. 392-405.

Rubin, David C. (ed.). 1986. Autobiographical memory. Cambridge.

Rubin, David C., and Alan D. Baddeley. 1989. Telescoping is not time compression: A Model of the Dating of Autobiographical Events, Memory & Cognition, 17,

Rustemeyer, Anitra. 1977. Measuring Interviewer Performance in Mock Interviews. Proceedings of the American Statistical Association (Social Statistics Section), pp. 341-346.

Schanz, Volker, and Peter Schmidt. 1984. Interviewsituation, Interviewermerkmale und Reaktionen von Befragten im Interview: eine multivariate Analyse. In Mayer, Karl U., and Peter Schmidt (eds.): 1984. Allgemeine Bevölkerungsumfrage der Sozialwissenschaften. Frankfurt (Main)/New York: Campus, pp. 72-113.

Schuman, H., and J. Converse. 1971. The Effects of Black and White Interviewers on Black Responses in 1968. Public Opinion Quarterly, 47, pp. 44-68.

Singer, Eleanor, Martin R. Frankel, and Marc B. Glassman. 1983. The Effect of Interviewer Characteristics and Expectations on Response. Public Opinion Quarterly, 47, pp. 68-83.

Singer, Eleanor, Luane Kohnke-Aguirre. 1979. Interviewer Expectations Effects: A Replication and Extension. Public Opinion Quarterly, 43, pp. 245-260.

Smith, T. W. 1981. Can We Have Confidence in Confidence? Revisited. In D. Johnston (ed.): Measurement of Subjective Phenomena. Washington, D.C.: Government Printing Office.

Sørensen, Aage B. 1991. Merton and Methodology. Contemporary Sociology, 20, pp. 516-519.

Sudman, Seymour und Norman M. Bradburn. 1973. Effects of Time and Memory Factors on Response in Surveys. Journal of the American Statistical Association, 68.

Sudman, Seymour und Norman M. Bradburn. 1974. Response Effects in Surveys: A Review and Synthesis. Chicago: Aldine.

Tölke, Angelika. 1980. Zuverlässigkeit retrospektiver Verlaufsdaten - Qualitative Ergebnisse einer Nachbefragung. Sonderforschungsbereich 3 "Mikroanalytische Grundlagen der Gesellschaftspolitik" der J.W.Goethe-Universität , Arbeitspapier Nr. 30, Frankfurt/Mannheim.

Tölke, Angelika. 1989. Möglichkeiten und Grenzen einer Edition bei retrospektiven Verlaufsdaten, in Karl Ulrich Mayer und Erika Brückner (eds.): Lebensverläufe und Wohlfahrtsentwicklung, pp. 173-226.

Tucker, Clyde. 1983. Interviewer Effects in Telephone Surveys. Public Opinion Quarterly, 47, pp. 84-95.

Van der Zouwen, J., and Wil Dijkstra. 1988. Types of Inadequate Interviewer Behavior in Survey Interviews: the Causes and Effects. Bulletin de Methodologie Sociologique, 18, pp. 5-20.

Voges, Wolfgang, und Michael Zwick. 1991. Die Bremer Stichprobe von Sozialhilfeakten: Chancen und Möglichkeiten für die empirische Sozialforschung. Zeitschrift für Soziologie 20(1), pp. 77-81.

Wagner, Michael. 1989. Korrekturen von Wohnverlaufsdaten: Ein Beispiel für die Aufbereitung von Retrospektivdaten. In Karl Ulrich Mayer und Erika Brückner (eds.): Lebensverläufe und Wohlfahrtsentwicklung, Teil I, pp. 239-246.

Weber, Max. 1949. The Methodology of the Social Sciences. New York: The Free Press.

Weiss, Carol H. 1975. Interviewing in Evaluation Research. In Elmer L. Struening and Marcia Guttentag (eds.): Handbook of Evaluation Research. Vol. 1. Beverly Hills, Ca.: Sage, pp. 355-396.

Witte, James C., and Arne L. Kalleberg. 1992. Vocational Training and Job Quality: Career Attainment in the German Labor Market. Paper presented at the 1992 Annual Meeting of the American Sociological Association, Pittsburgh, Pennsylvania.

Woolgar, Steve. 1983. Irony in the Social Study of Science. In Karin D. Knorr-Cetina, and Michael Mulkay (eds.): Science Observed. London: Sage, pp. 239-266.

I. Reihe STUDIEN UND BERICHTE des Max-Planck-Instituts für Bildungsforschung

Im Buchhandel erhältliche Bände (Vertrieb: edition sigma, Berlin)

61 Alexandra M. Freund
Die Selbstdefinition alter Menschen.
Inhalt, Struktur und Funktion.
251 S. Erschienen 1995.
ISBN 3-89404-808-5

60 Klaus Schömann
The Dynamics of Labor Earnings over the Life Course.
A Comparative and Longitudinal Analysis of Germany and Poland.
188 S. Erschienen 1994.
ISBN 3-89404-807-7

59 Frieder R. Lang
Die Gestaltung informeller Hilfebeziehungen im hohen Alter – Die Rolle von Elternschaft und Kinderlosigkeit.
Eine empirische Studie zur sozialen Unterstützung und deren Effekt auf die erlebte soziale Einbindung.
177 S. Erschienen 1994.
ISBN 3-89404-806-9

58 Ralf Th. Krampe
Maintaining Excellence.
Cognitive-Motor Performance in Pianists Differing in Age and Skill Level.
194 S. Erschienen 1994.
ISBN 3-89404-805-0

57 Ulrich Mayr
Age-Based Performance Limitations in Figural Transformations.
The Effect of Task Complexity and Practice.
172 S. Erschienen 1993.
ISBN 3-89404-804-2

56 Marc Szydlik
Arbeitseinkommen und Arbeitsstrukturen.
Eine Analyse für die Bundesrepublik Deutschland und die Deutsche Demokratische Republik.
255 S. Erschienen 1993.
ISBN 3-89404-803-4

55 Bernd Schellhas
Die Entwicklung der Ängstlichkeit in Kindheit und Jugend.
Befunde einer Längsschnittstudie über die Bedeutung der Ängstlichkeit für die Entwicklung der Kognition und des Schulerfolgs.
205 S. Erschienen 1993.
ISBN 3-89404-802-6

54 Falk Fabich
Forschungsfeld Schule: Wissenschaftsfreiheit, Individualisierung und Persönlichkeitsrechte.
Ein Beitrag zur Geschichte sozialwissenschaftlicher Forschung.
235 S. Erschienen 1993.
ISBN 3-89404-801-8

53 Helmut Köhler
Bildungsbeteiligung und Sozialstruktur in der Bundesrepublik.
Zu Stabilität und Wandel der Ungleichheit von Bildungschancen.
133 S. Erschienen 1992.
ISBN 3-89404-800-X

52 Ulman Lindenberger
Aging, Professional Expertise, and Cognitive Plasticity.
The Sample Case of Imagery-Based Memory Functioning in Expert Graphic Designers.
130 S. Erschienen 1991.
ISBN 3-608-98257-4

51 Volker Hofmann
Die Entwicklung depressiver Reaktionen in Kindheit und Jugend.
Eine entwicklungspsychopathologische Längsschnittuntersuchung.
197 S. Erschienen 1991.
ISBN 3-608-98256-6

50 Georgios Papastefanou
Familiengründung im Lebensverlauf.
Eine empirische Analyse sozialstruktureller Bedingungen der Familiengründung bei den Kohorten 1929–31, 1939–41 und 1949–51.
185 S. Erschienen 1990.
ISBN 3-608-98255-8

edition sigma
Heimstraße 14 D-10965 Berlin
Tel. 030 / 693 43 96 Fax 030 / 694 62 30

Ältere Bände (Nr. 1–42) nur noch beim Max-Planck-Institut für Bildungforschung erhältlich

I. Reihe STUDIEN UND BERICHTE des Max-Planck-Instituts für Bildungsforschung

Im Buchhandel erhältliche Bände (Vertrieb: edition sigma, Berlin)

49 Jutta Allmendinger
Career Mobility Dynamics.
A Comparative Analysis of the United States, Norway, and West Germany.
169 S. Erschienen 1989.
ISBN 3-608-98254-X

48 Doris Sowarka
Weisheit im Kontext von Person, Situation und Handlung.
Eine empirische Untersuchung alltagspsychologischer Konzepte alter Menschen.
275 S. Erschienen 1989.
ISBN 3-608-98253-1

47 Ursula M. Staudinger
The Study of Live Review.
An Approach to the Investigation of Intellectual Development Across the Life Span.
211 S. Erschienen 1989.
ISBN 3-608-98252-3

46 Detlef Oesterreich
Die Berufswahlentscheidung von jungen Lehrern.
115 S. Erschienen 1987.
ISBN 3-608-98251-5

45 Hans-Peter Füssel
Elternrecht und Schule.
Ein Beitrag zum Umfang des Elternrechts in der Schule für Lernbehinderte.
501 S. Erschienen 1987.
ISBN 3-608-98249-3

44 Diether Hopf
Herkunft und Schulbesuch ausländischer Kinder.
Eine Untersuchung am Beispiel griechischer Schüler.
114 S. Erschienen 1987.
ISBN 3-608-98248-5

43 Eberhard Schröder
Entwicklungssequenzen konkreter Operationen.
Eine empirische Untersuchung individueller Entwicklungsverläufe der Kognition.
112 S. Erschienen 1986.
ISBN 3-608-98247-7

edition sigma
Heimstraße 14 D-10965 Berlin
Tel. 030 / 693 43 96 Fax 030 / 694 62 30

Ältere Bände (Nr. 1–42) nur noch beim Max-Planck-Institut für Bildungforschung erhältlich

II. Reihe MATERIALIEN AUS DER BILDUNGSFORSCHUNG

Beim Max-Planck-Institut für Bildungsforschung erhältliche Bände (nicht über den Buchhandel beziehbar)

50 Hannah Brückner
Surveys Don't Lie, People Do?
An Analysis of Data Quality in a Retrospective Life Course Study.
86 S. Erschienen 1995.
ISBN 3-87985-042-9 DM 7,–

49 Todd D. Little, Gabriele Oettingen, and Paul B. Baltes
The Revised Control, Agency, and Means-ends Interview (CAMI).
A Multi-Cultural Validity Assessment Using Mean and Covariance Structures (MACS) Analyses.
97 S. Erschienen 1995.
ISBN 3-87985-041-0 DM 8,–

48 Hannah Brückner und Karl Ulrich Mayer
Lebensverläufe und gesellschaftlicher Wandel.
Konzeption, Design und Methodik der Erhebung von Lebensverläufen der Geburtsjahrgänge 1954–1956 und 1959–1961.
Teil I, Teil II, Teil III.
169 S., 224 S., 213 S.
Erschienen 1995.
ISBN 3-87985-039-9 DM 48,–

47 Jochen Fuchs
Die bundesdeutschen UNESCO-Projekt-Schulen und ihre internationalen Kontakte und Aktivitäten.
57 S. Erschienen 1995.
ISBN 3-87985-038-0 DM 7,–

46 Ursula M. Staudinger, Jacqui Smith und Paul B. Baltes
Handbuch zur Erfassung von weisheitsbezogenem Wissen.
89 S. Deutsche Ausgabe DM 10,–
Manual for the Assessment of Wisdom-Related Knowledge.
83 S. Englische Ausgabe DM 10,–
Erschienen 1994.
ISBN 3-87985-037-2

45 Jochen Fuchs
Internationale Kontakte im schulischen Sektor.
Zur Entwicklung und Situation des Schüleraustausches sowie von Schulpartnerschaften in der BRD.
174 S. Erschienen 1993.
ISBN 3-87985-035-6 DM 19,–

44 Erika Brückner
Lebensverläufe und gesellschaftlicher Wandel.
Konzeption, Design und Methodik der Erhebung von Lebensverläufen der Geburtsjahrgänge 1919–1921.
Teil I, Teil II, Teil III, Teil IV, Teil V.
235 S., 380 S., 200 S., 230 S., 141 S.
Erschienen 1993.
ISBN 3-87985-033-X DM 84,–

43 Ernst-H. Hoff und Hans-Uwe Hohner
Methoden zur Erfassung von Kontrollbewußtsein.
Textteil; Anhang.
99 S. und 178 S. Erschienen 1992.
ISBN 3-87985-032-1 DM 25,–

42 Michael Corsten und Wolfgang Lempert
Moralische Dimensionen der Arbeitssphäre.
Literaturbericht, Fallstudien und Bedingungsanalysen zum betrieblichen und beruflichen Handeln und Lernen.
367 S. Erschienen 1992.
ISBN 3-87985-031-3 DM 20,–

41 Armin Triebel
Zwei Klassen und die Vielfalt des Konsums.
Haushaltsbudgetierung bei abhängig Erwerbstätigen in Deutschland im ersten Drittel des 20. Jahrhunderts.
Teil I, Teil II.
416 S., 383 S. Erschienen 1991.
ISBN 3-87985-030-5 DM 48,–

40 Hans-Peter Füssel und Achim Leschinsky (Hrsg.)
Reform der Schulverfassung.
Wieviel Freiheit braucht die Schule?
Wieviel Freiheit verträgt die Schule?
117 S. Erschienen 1991.
ISBN 3-87985-029-1 DM 13,–

39 Gundel Schümer
Medieneinsatz im Unterricht.
Bericht über Ziel, Anlage und Durchführung einer Umfrage in allgemeinbildenden Schulen.
230 S. Erschienen 1991.
ISBN 3-87985-025-9 DM 24,–

II. Reihe MATERIALIEN AUS DER BILDUNGSFORSCHUNG (Fortsetzung)

38 Clemens Tesch-Römer
Identitätsprojekte und Identitätstransformationen im mittleren Erwachsenenalter.
312 S. Erschienen 1990.
ISBN 3-87985-026-7 DM 25,–

37 Helmut Köhler
Neue Entwicklungen des relativen Schul- und Hochschulbesuchs.
Eine Analyse der Daten für 1975 bis 1978.
138 S. Erschienen 1990.
ISBN 3-87985-024-0 DM 10,–

36 Wilfried Spang und Wolfgang Lempert
Analyse moralischer Argumentationen.
Beschreibung eines Auswertungsverfahrens.
Textteil: Grundlagen, Prozeduren, Evaluation.
Anhang: Interviewleitfaden, Tonbandtranskript und Auswertungsbeispiele.
102 und 191 S. Erschienen 1989. DM 29,–

35 Karl Ulrich Mayer und Erika Brückner
Lebensverläufe und Wohlfahrtsentwicklung.
Konzeption, Design und Methodik der Erhebung von Lebensverläufen der Geburtsjahrgänge 1929–1931, 1939–1941, 1949–1951.
Teil I, Teil II, Teil III.
261 S., unpaginiert, 175 S.
Erschienen 1989. DM 39,–

34 Christoph Droß und Wolfgang Lempert
Untersuchungen zur Sozialisation in der Arbeit 1977 bis 1988.
Ein Literaturbericht.
204 S. Erschienen 1988. DM 12,–

32 Friedrich Edding (Hrsg.)
Bildung durch Wissenschaft in neben- und nachberuflichen Studien.
Tagungsbericht.
157 S. Erschienen 1988. DM 11,–

31 Ellen A. Skinner, Michael Chapman and Paul B. Baltes
The Control, Agency, and Means-Ends Beliefs Interview.
A New Measure of Perceived Control in Children (School Domain).
Ein neues Meßinstrument für Kontrollüberzeugungen bei Kindern (Bereich Schule).
54 S. Erschienen 1988. DM 9,–

29 Ulrich Trommer
Aufwendungen für Forschung und Entwicklung in der Bundesrepublik Deutschland 1965 bis 1983.
Theoretische und empirisch-statistische Probleme.
321 S. Erschienen 1987. DM 32,–

28 Ingeborg Tölke
Ein dynamisches Schätzverfahren für latente Variablen in Zeitreihenanalysen.
202 S. Erschienen 1986. DM 17,–

Die nicht aufgeführten Bände sind vergriffen, bzw. nur noch in Restexemplaren erhältlich.

III. Einzelpublikationen

Beim Max-Planck-Institut für Bildungsforschung erhältliche Titel (nicht über den Buchhandel beziehbar)

Max-Planck-Institut für Bildungsforschung (Hrsg.)
Bekenntnis und Dienst.
Reden zum 80. Geburtstag von Dietrich Goldschmidt.
96 S. Berlin: Max-Planck-Institut für Bildungsforschung, 1995.
ISBN 3-87985-040-2

Max-Planck-Institut für Bildungsforschung (Hrsg.)
Abschied von Hellmut Becker.
Reden auf der Trauerfeier am 18. Januar 1994.
47 S. Berlin: Max-Planck-Institut für Bildungsforschung, 1994.
ISBN 3-87985-036-4

Max-Planck-Institut für Bildungsforschung (Hrsg.)
Bildungsforschung und Bildungspolitik.
Reden zum 80. Geburtstag von Hellmut Becker.
98 S. Berlin: Max-Planck-Institut für Bildungsforschung, 1993.
ISBN 3-87985-034-8

Wolfgang Schneider and Wolfgang Edelstein (Eds.)
Inventory of European Longitudinal Studies in the Behavioral and Medical Sciences.
A Project Supported by the European Science Foundation.
557 S. Munich: Max Planck Institute for Psychological Research, and Berlin: Max Planck Institute for Human Development and Education, 1990.
ISBN 3-87985-028-3
DM 58,–

Max-Planck-Institut für Bildungsforschung (Hrsg.)
Entwicklung und Lernen.
Beiträge zum Symposium anläßlich des 60. Geburtstages von Wolfgang Edelstein.
98 S. Berlin: Max-Planck-Institut für Bildungsforschung, 1990.
ISBN 3-87985-023-2

Max-Planck-Institut für Bildungsforschung (Hrsg.)
Normative Voraussetzungen und ethische Implikationen sozialwissenschaftlicher Forschung.
Beiträge zum Symposium anläßlich des 75. Geburtstages von Dietrich Goldschmidt.
108 S. Berlin: Max-Planck-Institut für Bildungsforschung, 1990.
ISBN 3-87985-027-5

Max-Planck-Institut für Bildungsforschung (Hrsg.)
25 Jahre Max-Planck-Institut für Bildungsforschung.
Festvorträge.
48 S. Berlin: Max-Planck-Institut für Bildungsforschung, 1989.

Friedrich Edding
Mein Leben mit der Politik.
126 S. Berlin: Max-Planck-Institut für Bildungsforschung, 1989.

Max-Planck-Institut für Bildungsforschung (Hrsg.)
Gewerbliche Unternehmen als Bildungsträger.
Beiträge zum Symposium anläßlich des 80. Geburtstages von Friedrich Edding.
126 S. Berlin: Max-Planck-Institut für Bildungsforschung, 1989.

Weitere Schriftenreihen aus dem Max-Planck-Institut für Bildungsforschung (nicht über den Buchhandel erhältlich)

- Beiträge aus dem Forschungsbereich Entwicklung und Sozialisation (bitte Liste der Veröffentlichungen anfordern)
- Beiträge aus dem Forschungsbereich Schule und Unterricht (bitte Liste der Veröffentlichungen anfordern)
- Literatur-Informationen aus der Bildungsforschung (monatliche Neuerwerbungen der Bibliothek; Abonnement DM 60,–/Jahr)

IV. Buchveröffentlichungen bei Verlagen (nach dem Erscheinungsjahr geordnet, nur lieferbare Titel; nur über den Buchhandel zu beziehen)

Lothar Krappmann und Hans Oswald
Alltag der Schulkinder.
Beobachtungen und Analysen von Interaktionen und Sozialbeziehungen.
224 S. Weinheim/München: Juventa, 1995.

Freya Dittmann-Kohli
Das persönliche Sinnsystem.
Ein Vergleich zwischen frühem und spätem Erwachsenenalter.
402 S. Göttingen/Bern/Toronto/Seattle: Hogrefe, 1995.

Hartmut Zeiher und Helga Zeiher
Orte und Zeiten der Kinder.
Soziales Leben im Alltag von Großstadtkindern.
223 S. Weinheim/München: Juventa, 1994.

Christiane Lange-Küttner
Gestalt und Konstruktion.
Die Entwicklung der grafischen Kompetenz beim Kind.
242 S. Bern/Toronto: Huber, 1994.

Jutta Allmendinger
Lebensverlauf und Sozialpolitik.
Die Ungleichheit von Mann und Frau und ihr öffentlicher Ertrag.
302 S. Frankfurt a. M./New York: Campus, 1994.

Wolfgang Lauterbach
Berufsverläufe von Frauen.
Erwerbstätigkeit, Unterbrechung und Wiedereintritt.
289 S. Frankfurt a. M./New York: Campus, 1994.

Arbeitsgruppe Bildungsbericht am Max-Planck-Institut für Bildungsforschung
Das Bildungswesen in der Bundesrepublik Deutschland.
Strukturen und Entwicklungen im Überblick.
843 S. Reinbek: Rowohlt, 1994 (4., vollständig überarbeitete und erweiterte Neuausgabe).

Hellmut Becker und Gerhard Kluchert
Die Bildung der Nation.
Schule, Gesellschaft und Politik vom Kaiserreich zur Weimarer Republik.
538 S. Stuttgart: Klett-Cotta, 1993.

Rolf Becker
Staatsexpansion und Karrierechancen.
Berufsverläufe im öffentlichen Dienst und in der Privatwirtschaft.
303 S. Frankfurt a.M./New York: Campus, 1993.

Wolfgang Edelstein und Siegfried Hoppe-Graff (Hrsg.)
Die Konstruktion kognitiver Strukturen. Perspektiven einer konstruktivistischen Entwicklungspsychologie.
328 S. Bern/Stuttgart/Toronto: Huber, 1993.

Wolfgang Edelstein, Gertrud Nunner-Winkler und Gil Noam (Hrsg.)
Moral und Person.
418 S. Frankfurt a.M.: Suhrkamp, 1993.

Lothar Lappe
Berufsperspektiven junger Facharbeiter.
Eine qualitative Längsschnittanalyse zum Kernbereich westdeutscher Industriearbeit.
394 S. Frankfurt a.M./New York: Campus, 1993.

Detlef Oesterreich
Autoritäre Persönlichkeit und Gesellschaftsordnung.
Der Stellenwert psychischer Faktoren für politische Einstellungen – eine empirische Untersuchung von Jugendlichen in Ost und West.
243 S. Weinheim/München: Juventa, 1993.

Marianne Müller-Brettel
Bibliographie Friedensforschung und Friedenspolitik:
Der Beitrag der Psychologie 1900–1991.
(Deutsch/Englisch)
383 S. München/London/New York/Paris: Saur, 1993.

Paul B. Baltes und Jürgen Mittelstraß (Hrsg.)
Zukunft des Alterns und gesellschaftliche Entwicklung.
(= Forschungsberichte der Akademie der Wissenschaften zu Berlin, 5.)
814 S. Berlin/New York: De Gruyter, 1992.

Matthias Grundmann
Familienstruktur und Lebensverlauf.
Historische und gesellschaftliche Bedingungen individueller Entwicklung.
226 S. Frankfurt a.M./New York: Campus, 1992.

IV. Buchveröffentlichungen bei Verlagen (Fortsetzung)

Karl Ulrich Mayer (Hrsg.)
Generationsdynamik in der Forschung.
245 S. Frankfurt a.M./New York: Campus, 1992.

Erika M. Hoerning
Zwischen den Fronten.
Berliner Grenzgänger und Grenzhändler 1948–1961.
266 S. Köln/Weimar/Wien: Böhlau, 1992.

Ernst-H. Hoff
Arbeit, Freizeit und Persönlichkeit.
Wissenschaftliche und alltägliche Vorstellungsmuster.
238 S. Heidelberg: Asanger Verlag, 1992 (2. überarbeitete und aktualisierte Auflage).

Erika M. Hoerning
Biographieforschung und Erwachsenenbildung.
223 S. Bad Heilbrunn: Klinkhardt, 1991.

Max-Planck-Institut für Bildungsforschung
Traditions et transformations.
Le système d'éducation en République fédérale d'Allemagne.
341 S. Paris: Economica, 1991.

Dietrich Goldschmidt
Die gesellschaftliche Herausforderung der Universität.
Historische Analysen, internationale Vergleiche, globale Perspektiven.
297 S. Weinheim: Deutscher Studien Verlag, 1991.

Uwe Henning und Achim Leschinsky (Hrsg.)
Enttäuschung und Widerspruch.
Die konservative Position Eduard Sprangers im Nationalsozialismus. Analysen – Texte – Dokumente.
213 S. Weinheim: Deutscher Studien Verlag, 1991.

Ernst-H. Hoff, Wolfgang Lempert und Lothar Lappe
Persönlichkeitsentwicklung in Facharbeiterbiographien.
282 S. Bern/Stuttgart/Toronto: Huber, 1991.

Karl Ulrich Mayer, Jutta Allmendinger und Johannes Huinink (Hrsg.)
Vom Regen in die Traufe: Frauen zwischen Beruf und Familie.
483 S. Frankfurt a.M./New York: Campus, 1991.

Maria von Salisch
Kinderfreundschaften.
Emotionale Kommunikation im Konflikt.
153 S. Göttingen/Toronto/Zürich: Hogrefe, 1991.

Paul B. Baltes and Margret M. Baltes (Eds.)
Successful Aging: Perspectives from the Behavioral Sciences.
397 pp. Cambridge: Cambridge University Press, 1990.

Paul B. Baltes, David L. Featherman and Richard M. Lerner (Eds.)
Life-Span Development and Behavior.
368 pp. Vol. 10. Hillsdale, N.J.: Erlbaum, 1990.

Achim Leschinsky and Karl Ulrich Mayer (Eds.)
The Comprehensive School Experiment Revisited: Evidence from Western Europe.
211 pp. Frankfurt a.M./Bern/New York/Paris: Lang 1990.

Karl Ulrich Mayer (Hrsg.)
Lebensverläufe und sozialer Wandel.
467 S. Opladen: Westdeutscher Verlag, 1990.
(= Sonderheft 31 der KZfSS).

Karl Ulrich Mayer and Nancy Brandon Tuma (Eds.)
Event History Analysis in Life Course Research.
320 pp. Madison, Wis.: The University of Wisconsin Press, 1990.

Hans J. Nissen, Peter Damerow und Robert K. Englund
Frühe Schrift und Techniken der Wirtschaftsverwaltung im alten Vorderen Orient.
Informationsspeicherung und -verarbeitung vor 5000 Jahren.
Katalog zur gleichnamigen Ausstellung Berlin-Charlottenburg, Mai–Juli 1990.
222 S. Bad Salzdetfurth: Franzbecker, 1990.
(2. Aufl. 1991).

Peter Alheit und Erika M. Hoerning (Hrsg.)
Biographisches Wissen.
Beiträge zu einer Theorie lebensgeschichtlicher Erfahrung.
284 S. Frankfurt a.M./New York: Campus, 1989.

Arbeitsgruppe am Max-Planck-Institut für Bildungsforschung
Das Bildungswesen in der Bundesrepublik Deutschland.
Ein Überblick für Eltern, Lehrer und Schüler.
Japanische Ausgabe: 348 S. Tokyo: Toshindo Publishing Co. Ltd., 1989.

IV. Buchveröffentlichungen bei Verlagen (Fortsetzung)

Hans-Peter Blossfeld
Kohortendifferenzierung und Karriereprozeß.
Eine Längsschnittstudie über die Veränderung der Bildungs- und Berufschancen im Lebenslauf.
185 S. Frankfurt a.M./New York: Campus, 1989.

Hans-Peter Blossfeld, Alfred Hamerle and Karl Ulrich Mayer
Event History Analysis.
Statistical Theory and Application in the Social Sciences.
297 pp. Hillsdale, N.J.: Erlbaum, 1989.

Erika M. Hoerning und Hans Tietgens (Hrsg.)
Erwachsenenbildung: Interaktion mit der Wirklichkeit.
200 S. Bad Heilbrunn: Klinkhardt, 1989.

Johannes Huinink
Mehrebenensystem-Modelle in den Sozialwissenschaften.
292 S. Wiesbaden: Deutscher Universitäts-Verlag, 1989.

Kurt Kreppner and Richard M. Lerner (Eds.)
Family Systems and Life-Span Development.
416 pp. Hillsdale, N.J.: Erlbaum, 1989.

Bernhard Schmitz
Einführung in die Zeitreihenanalyse.
Modelle, Softwarebeschreibung, Anwendungen.
235 S. Bern/Stuttgart/Toronto: Huber, 1989.

Eberhard Schröder
Vom konkreten zum formalen Denken.
Individuelle Entwicklungsverläufe von der Kindheit zum Jugendalter.
328 S. Bern/Stuttgart/Toronto: Huber, 1989.

Michael Wagner
Räumliche Mobilität im Lebensverlauf.
Eine empirische Untersuchung sozialer Bedingungen der Migration.
226 S. Stuttgart: Enke, 1989.

Paul B. Baltes, David L. Featherman and Richard M. Lerner (Eds.)
Life-Span Development and Behavior.
338 pp. Vol. 9. Hillsdale, N.J.: Erlbaum, 1988.

Paul B. Baltes, David L. Featherman and Richard M. Lerner (Eds.)
Life-Span Development and Behavior.
337 pp. Vol. 8. Hillsdale, N.J.: Erlbaum, 1988.

Lothar Krappmann
Soziologische Dimensionen der Identität.
Strukturelle Bedingungen für die Teilnahme an Interaktionsprozessen.
231 S. Stuttgart: Klett-Cotta, 7. Aufl., 1988
(= Standardwerke der Psychologie).

Detlef Oesterreich
Lehrerkooperation und Lehrersozialisation.
159 S. Weinheim: Deutscher Studien Verlag, 1988.

Michael Bochow und Hans Joas
Wissenschaft und Karriere.
Der berufliche Verbleib des akademischen Mittelbaus.
172 und 37 S. Frankfurt a.M./New York: Campus, 1987.

Hans-Uwe Hohner
Kontrollbewußtsein und berufliches Handeln.
Motivationale und identitätsbezogene Funktionen subjektiver Kontrollkonzepte.
201 S. Bern/Stuttgart/Toronto: Huber, 1987.

Bernhard Schmitz
Zeitreihenanalyse in der Psychologie.
Verfahren zur Veränderungsmesung und Prozeßdiagnostik.
304 S. Weinheim/Basel: Deutscher Studien Verlag/ Beltz, 1987.

Margret M. Baltes and Paul B. Baltes (Eds.)
The Psychology of Control and Aging.
415 pp. Hillsdale, N.J.: Erlbaum, 1986.

Paul B. Baltes, David L. Featherman and Richard M. Lerner (Eds.)
Life-Span Development and Behavior.
334 pp. Vol. 7. Hillsdale, N.J.: Erlbaum, 1986.

Hans-Peter Blossfeld, Alfred Hamerle und Karl Ulrich Mayer
Ereignisanalyse.
Statistische Theorie und Anwendung in den Wirtschafts- und Sozialwissenschaften.
290 S. Frankfurt a.M./New York: Campus, 1986.

Axel Funke, Dirk Hartung, Beate Krais und Reinhard Nuthmann
Karrieren außer der Reihe.
Bildungswege und Berufserfolge von Stipendiaten der gewerkschaftlichen Studienförderung.
256 S. Köln: Bund, 1986.

IV. Buchveröffentlichungen bei Verlagen (Fortsetzung)

Ernst-H. Hoff, Lothar Lappe und Wolfgang Lempert (Hrsg.)
Arbeitsbiographie und Persönlichkeitsentwicklung.
288 S. Bern/Stuttgart/Toronto: Huber, 1986.

Klaus Hüfner, Jens Naumann, Helmut Köhler und Gottfried Pfeffer
Hochkonjunktur und Flaute: Bildungspolitik in der Bundesrepublik Deutschland 1967–1980.
361 S. Stuttgart: Klett-Cotta, 1986.

Jürgen Staupe
Parlamentsvorbehalt und Delegationsbefugnis.
Zur „Wesentlichkeitstheorie" und zur Reichweite legislativer Regelungskompetenz, insbesondere im Schulrecht.
419 S. Berlin: Duncker & Humblot, 1986.

Hans-Peter Blossfeld
Bildungsexpansion und Berufschancen.
Empirische Analysen zur Lage der Berufsanfänger in der Bundesrepublik.
191 S. Frankfurt a.M./New York: Campus, 1985.

Christel Hopf, Knut Nevermann und Ingrid Schmidt
Wie kamen die Nationalsozialisten an die Macht.
Eine empirische Analyse von Deutungen im Unterricht.
344 S. Frankfurt a.M./New York: Campus, 1985.

John R. Nesselroade and Alexander von Eye (Eds.)
Individual Development and Social Change: Explanatory Analysis.
380 pp. New York: Academic Press, 1985.

Michael Jenne
Music, Communication, Ideology.
185 pp. Princeton, N.J.: Birch Tree Group Ltd., 1984.

Gero Lenhardt
Schule und bürokratische Rationalität.
282 S. Frankfurt a.M.: Suhrkamp, 1984.

Achim Leschinsky und Peter Martin Roeder
Schule im historischen Prozeß.
Zum Wechselverhältnis von institutioneller Erziehung und gesellschaftlicher Entwicklung.
545 S. Frankfurt a.M./Berlin/Wien: Ullstein, 1983.

Max Planck Institute for Human Development and Education
Between Elite and Mass Education.
Education in the Federal Republic of Germany.
348 pp. Albany: State University of New York Press, 1983.

Margit Osterloh
Handlungsspielräume und Informationsverarbeitung.
369 S. Bern/Stuttgart/Toronto: Huber, 1983.

Knut Nevermann
Der Schulleiter.
Juristische und historische Aspekte zum Verhältnis von Bürokratie und Pädagogik.
314 S. Stuttgart: Klett-Cotta, 1982.

Gerd Sattler
Englischunterricht im FEGA-Modell.
Eine empirische Untersuchung über inhaltliche und methodische Differenzierung an Gesamtschulen.
355 S. Stuttgart: Klett-Cotta, 1981.

Christel Hopf, Knut Nevermann und Ingo Richter
Schulaufsicht und Schule.
Eine empirische Analyse der administrativen Bedingungen schulischer Erziehung.
428 S. Stuttgart: Klett-Cotta, 1980.

Diether Hopf
Mathematikunterricht.
Eine empirische Untersuchung zur Didaktik und Unterrichtsmethode in der 7. Klasse des Gymnasiums.
251 S. Stuttgart: Klett-Cotta, 1980.

Max-Planck-Institut für Bildungsforschung
Projektgruppe Bildungsbericht (Hrsg.)
Bildung in der Bundesrepublik Deutschland.
Daten und Analysen.
Bd. 1: Entwicklungen seit 1950.
Bd. 2: Gegenwärtige Probleme.
1404 S. Stuttgart: Klett-Cotta, 1980.

Dietrich Goldschmidt und Peter Martin Roeder (Hrsg.)
Alternative Schulen?
Gestalt und Funktion nichtstaatlicher Schulen im Rahmen öffentlicher Bildungssysteme.
623 S. Stuttgart: Klett-Cotta, 1979.