

GESIS

ZUMA

NACHRICHTEN

54

Mai 2004

Zentrum für Umfragen, Methoden und Analysen (ZUMA)

ZUMA ist Mitglied der Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen e.V. (GESIS). Die GESIS ist eine Einrichtung der *Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz* (WGL) (www.wgl.de).

Vorsitzender des Trägervereins ZUMA e.V.: Prof. Dr. Jan van Deth

Direktor: Prof. Dr. Peter Ph. Mohler

Hausanschrift

B 2, 1
68 159 Mannheim

Postanschrift

Postfach 12 21 55
68 072 Mannheim

Telefon 0621/1246 - 0

Fax 0621/1246 - 100

Redaktion: 0621/1246 - 228

E-Mail porst@zuma-mannheim.de

Internet www.gesis.org/zuma ZUMA

www.gesis.org/ GESIS

Herausgeber: Prof. Dr. Ingwer Borg

Redaktion: Rolf Porst

ISSN 0941-1670 28. Jahrgang

Die ZUMA-Nachrichten veröffentlichen - neben Nachrichten des Instituts - Artikel, die von Interesse für die empirische Sozialforschung, insbesondere die Umfrageforschung, sind. Alle Artikel müssen einen methodischen Fokus haben oder zumindest methodische Aspekte stark betonen. Das Spektrum möglicher Artikel ist breit: Es reicht von Grundlagenforschung über angewandte Papiere bis zu Arbeiten, die einen praktisch-operativen Charakter haben. Die Artikel in den ZUMA-Nachrichten sollen für eine breite Leserschaft von Wissenschaftlern und Praktikern im Bereich der empirischen Sozialforschung verständlich sein. Alle Beiträge, die zur Veröffentlichung in den ZN eingereicht werden, werden von mindestens zwei unabhängigen Gutachtern blind begutachtet.

Namentlich gekennzeichnete Beiträge geben die Meinung der Autoren wieder. Der Nachdruck von Beiträgen ist nach Absprache möglich. Die ZUMA-Nachrichten erscheinen im Mai und November eines Jahres. Sie werden Interessenten auf Anforderung kostenlos zugesandt. Die ZUMA-Nachrichten finden Sie auch im WWW unter der Adresse: http://www.gesis.org/publikationen/zeitschriften/zuma_nachrichten/.

Druck: PrintArt GmbH Druckerei + Verlag, Dannstadt/Mannheim
Gedruckt auf chlorfrei gebleichtem Papier.

© ZUMA, Mannheim, 2004

Editorial	5
-----------------	---

FORSCHUNGSBERICHTE

Repräsentativität der Lebensverlaufsstudie Ost (Geburtskohorte 1971). Ein Vergleich ausgewählter Randverteilungen mit den Erhebungen des Mikrozensus. <i>Britta Matthes & Sigrid Wehner</i>	6
Selektives Antwortverhalten bei Fragen zum delinquenten Handeln. Eine empirische Studie über die Wirksamkeit der „sealed envelope technique“ bei selbst berichteter Delinquenz mit Daten des ALLBUS 2000. <i>Rolf Becker & Ralph Günther</i>	39
Kinder und Jugendliche als Befragte. Feldexperimente zum Antwortverhalten Minderjähriger. <i>Marek Fuchs</i>	60
Determinanten und Konsequenzen der Umfrageeinstellung. Bewertungsdimensionen unterschiedlicher Umfragesponsoren und die Antwortbereitschaft der Befragten. <i>Volker Stocké & Birgit Becker</i>	89
Computerunterstützte Inhaltsanalyse ohne Diktionär? Ein Praxistest. <i>Juliane Landmann & Cornelia Züll</i>	117
Erfahrungen mit Personennamen zur Bildung von Stichproben für Betriebsbefragungen. <i>Andreas Humpert</i>	141
Zur Umsetzung des Internationalen Sozioökonomischen Index des beruflichen Status (ISEI) mit den Mikrozensen ab 1996. <i>Bernhard Schimpl-Neimanns</i>	154

BUCHBESPRECHUNGEN

Sachs, Lothar: Angewandte Statistik, 11. Auflage, 2004.....	171
Wittenberg, Reinhard & Hans Cramer: Datenanalyse mit SPSS für Windows, Handbuch für computerunterstützte Datenanalyse Band IX, 3. neubearbeitete Auflage, 2003.....	173

ZUMA-PUBLIKATIONEN

ZUMA-Methodenberichte.....	178
ZUMA „How to-Reihe“	179

ZUMA-VERANSTALTUNGEN 2004

ZUMA-workshops: Zweites Halbjahr 2004	181
---	-----

NACHRICHTEN AUS DEM INSTITUT	182
---	-----

DURCHWAHL-RUFNUMMERN	183
-----------------------------------	-----

EDITORIAL

Mit der vorliegenden Ausgabe der ZUMA-Nachrichten ergeben sich zwei personelle Veränderungen in Herausgabe und Redaktion der Zeitschrift. Herausgeber ist ab jetzt INGWER BORG (bei ZUMA als wissenschaftlicher Leiter für den Bereich Wissensvermittlung und Beratung zuständig und zudem Professor für angewandte psychologische Methodik an der Universität Gießen), als Redakteur fungiert ROLF PORST (ansonsten bei ZUMA als Projektberater tätig).

PAUL LÜTTINGER, der über viele Jahre als Redakteur für die ZUMA-Nachrichten verantwortlich zeichnete, hat seine Arbeit an der Zeitschrift beendet. Wir danken ihm herzlich für die lange, erfolgreiche Tätigkeit!

Die personellen Veränderungen bei den Verantwortlichkeiten werden den Charakter der ZUMA-Nachrichten nicht verändern. Die ZUMA-Nachrichten sollen ihren in der Profession wohl-akzeptierten Mix aus Grundlagen- und angewandter Forschung im weiten Umfeld sozialwissenschaftlicher Umfragen behalten, inklusive gelegentlicher eher handwerklicher Papiere für den Praktiker. Schwerpunkt dabei bleibt grundsätzlich die Methodenforschung. Alle zur Veröffentlichung eingereichten Manuskripte werden, wie auch bisher schon, von zwei Gutachtern „blind“ begutachtet.

REPRÄSENTATIVITÄT DER LEBENSVERLAUFSSTUDIE OST (GEBURTSKOHORTE 1971)

Ein Vergleich ausgewählter Randverteilungen mit den Erhebungen des Mikrozensus

BRITTA MATTHES & SIGRID WEHNER

Die Lebensverlaufsstudie Ost (Geburtskohorte 1971) ist eine umfangreiche retrospektive Lebensverlaufserhebung, die im Rahmen des Teilprojektes „Ostdeutsche Lebensverläufe im Transformationsprozess“ am Max-Planck-Institut für Bildungsforschung, Berlin, realisiert wurde. Da der Zeitraum zwischen Stichprobenziehung (Oktober 1990) und Interview (Mai 1996 bis Januar 1998) relativ lang und der Anteil der Verweigerer ungewöhnlich hoch war, haben wir – um eventuelle Stichprobenverzerrungen analytisch zu überprüfen – ausgewählte Randverteilungen der realisierten Stichprobe der Lebensverlaufsstudie Ost (Geburtskohorte 1971) mit denen der Mikrozensusserhebungen aus den Jahren 1991, 1993, 1995 und 1996 verglichen. Dieser Vergleich ergab, dass in der Lebensverlaufsstudie Ost (Geburtskohorte 1971) – neben für solche Erhebungen typischen Verzerrungen – auch Abweichungen aufgetreten sind, die auf Schwierigkeiten bei der Recherche der aktuellen Adressen zurückgeführt werden können. Diese Abweichungen sind allerdings bei der Untersuchung von Wechselbeziehungen zwischen Variablen (und deren Veränderung in der Zeit) unwichtig, sofern der identifizierte Selektionsbias berücksichtigt wird.

The survey of the East German cohort 1971 is an elaborate retrospective life course study, collected within the context of the German Life History Study “East German Life Courses after Unification” at the Max Planck Institute for Human Development, Berlin. Due to the long period between sampling (October 1990) and interviewing (May 1996 to January 1998) and due to the unusually high number of refusals, selected marginal distributions of the East German birth cohort 1971 sample and the Mikrozensus sample (official statistical census in Germany) from 1991, 1993, 1995 and 1996 were

compared in order to identify any distortions in the 1971 sample. One main result from this comparison is that, apart from divergences characteristic of life course surveys, the survey of the East German birth cohort 1971 has deviations linked to problems in tracing current addresses. Provided the identified selection bias is taken into account, these deviations are not important for the analysis of interactions between variables (and their changes over time).

1. Einleitung¹

Bei der Erhebung der Lebensverläufe der DDR-Geburtskohorte 1971,² die im Rahmen der Lebensverlaufsstudie „Ostdeutsche Lebensverläufe im Transformationsprozess“ am Max-Planck-Institut für Bildungsforschung, Berlin, im Forschungsbereich „Bildung, Arbeit und gesellschaftliche Entwicklung“ unter der Leitung von Prof. Dr. Karl Ulrich Mayer durchgeführt wurde, konnte nur eine Ausschöpfungsquote³ von 48,0 Prozent erreicht werden (vgl. INFAS 1998). Insbesondere der Anteil der Zielpersonen, die eine Teilnahme an der Studie verweigerten, war mit 41,5 Prozent sehr hoch.⁴ Darüber hinaus lag, da der Zeitraum zwischen dem Zeitpunkt der Stichprobenziehung (Oktober 1990) und dem Interview (Mai 1996 bis Januar 1998) relativ lang war, der Anteil stichprobenneutraler Ausfälle (Adresse unbekannt, Wohnung unbewohnt, Zielperson unbekannt⁵, Zielperson neue Adresse, Zielperson verstorben) mit 32,1 Prozent – trotz verstärkter Anstrengungen bei der Recherche der aktuellen Adressen – deutlich über dem bei Lebensverlaufsuntersuchungen üblichen Umfang.⁶

1 Für wertvolle Hinweise und Kommentare bedanken wir uns herzlich bei unserem anonymen Gutachter.

2 Die Lebensverlaufsstudie „Ostdeutsche Lebensverläufe im Transformationsprozess (LV-Ost)“ für die Geburtskohorte 1971 wird im folgenden kurz LV-Ost 71 genannt.

3 Die Ausschöpfungsquote entspricht dabei dem Anteil der durchgeführten Interviews an der bereinigten Stichprobe (Bruttostichprobe abzüglich stichprobenneutraler Ausfälle).

4 Dabei fällt die hohe Differenz zwischen den Erhebungsmethoden ins Auge. Die Verweigerungen waren bei der telefonischen Befragung deutlich niedriger (38 Prozent) als bei der Face-to-Face-Erhebung (48 Prozent). Die von den Interviewern aufgenommenen Verweigerungsgründe zeigen, dass mangelndes Interesse der Zielpersonen und eine grundsätzliche Ablehnung der Interviewbereitschaft zu den wichtigsten Verweigerungsgründen zählen.

5 Die Annahme der Stichprobenneutralität der Ausfälle aufgrund Nichterreichbarkeit der Zielperson ist insbesondere dann zu hinterfragen, wenn regionale Mobilität eine wichtige Voraussetzung für bessere Bildungs- und Erwerbschancen beziehungsweise für beruflichen Erfolg ist.

6 Ausschließlich bei Face-to-Face-Erhebungen, die nur bei telefonischer Nichterreichbarkeit und auf Wunsch der Zielpersonen durchgeführt wurden, wurde der Bearbeitungsstatus „Zielperson

Untersuchungen belegen, dass die Verweigerung eines Interviews eher situativ bedingt ist, und sich die Verweigerer deshalb nur wenig von den Teilnehmenden, wohl aber von den Nichterreichbaren unterscheiden (Esser 1986: 38, De Leeuw 1992: 41-44). Bisher ist allerdings nur unzureichend analysiert worden, ob und welche Verzerrungen der ursprünglichen Stichprobe sich durch stichprobenneutrale Ausfälle ergeben. Unter der Annahme, dass insbesondere regional mobile Zielpersonen viel schwerer zu erreichen sind und dass regionale Mobilität sozial selektiv erfolgt, ist auch bei der hier betrachteten Lebensverlaufsstudie eine Stichprobenverzerrung durch „stichprobenneutrale“ Ausfälle nicht auszuschließen. Zwar wurden durch eine aufwändige Recherche der aktuellen Adressen auch regional mobile Zielpersonen befragt, jedoch ist davon auszugehen, dass regional immobile Zielpersonen (bzw. Zielpersonen, deren Eltern regional immobil waren), mit höherer Wahrscheinlichkeit recherchierbar sind.

Vor diesem Hintergrund erschien es uns angebracht, die Repräsentativität der realisierten Stichprobe der LV-Ost 71 analytisch zu überprüfen. Häufig wird die Notwendigkeit einer externen Validierung von Lebensverlaufsdaten diskutiert (Blossfeld 1987; Rendtel/Pötter 1993; Born/Erzberger 1999). Da sich die Repräsentativität einer Stichprobe jedoch auf die Übereinstimmung der multidimensionalen Verteilungen aller Merkmale bezieht, kann sie in der Regel empirisch nicht vollständig überprüft werden. In der sozialwissenschaftlichen Forschung wird sie deshalb zumeist analysiert, indem die multivariaten Randverteilungen vergleichbarer Merkmale einer Stichprobe denen in der amtlichen Statistik gegenübergestellt werden (Hartmann/ Schimpl-Neimanns 1992: 318; Rendtel/Pötter 1993: 260).⁷ Da es sich beim Mikrozensus um die Querschnittserhebung einer sehr großen Stichprobe handelt, bei der – zumindest für einen Teil der Variablen – Auskunftspflicht besteht, ist davon auszugehen, dass die Mikrozensusdaten bestimmte Merkmale der Grundgesamtheit exakter widerspiegeln als die Daten der Lebensverlaufsstudie. Um die Möglichkeiten und Grenzen eines solchen Validierungsversuchs auszuloten, werden im

unbekannt“ vergeben (15 Prozent). Telefonischen Befragung erscheinen deshalb vorteilhafter als Face-to-Face-Erhebungen, da bei der telefonischen Erhebung mobile Zielpersonen weit häufiger kontaktiert, die Telefonnummern der Zielpersonen durch andere Kontaktpersonen am Telefon direkt übermittelt beziehungsweise veränderte Telefonnummern selbst recherchiert werden können.

7 Diese Definition der Repräsentativität wird insbesondere deswegen kritisiert, weil auf der Basis der Verteilung sozialstruktureller Merkmale keine Erkenntnisse über den tatsächlichen Prozess der Nichtteilnahme gewonnen werden können (Schnell 1997: 199). Eine Alternative ist die Erhebung und Analyse von Nonresponse-Merkmalen (z.B. Erbslöh/Koch 1988). Da uns in der LV-Ost 71 jedoch keine Informationen über die Verweigerer und die Nichterreichbaren zur Verfügung stehen, ist eine solche Analyse leider nicht möglich. Allerdings können „die Ergebnisse von Stichprobenvergleichen als indirekte Analyse der Stichprobenausfälle interpretiert werden“ (Rendtel/Pötter 1993: 278).

Folgenden zunächst die grundlegenden Charakteristika der Lebensverlaufsstudie LV-Ost 71 und des Mikrozensus gegenübergestellt.

Streng genommen ist die Beurteilung der Stichprobengüte der LV-Ost 71 nur für den Zeitpunkt der Stichprobenziehung möglich, da sich die Grundgesamtheit z.B. durch regionale Mobilität ständig verändern kann. Das bedeutet, dass selbst bei vollständiger Zufallsauswahl für retrospektiv erfasste Daten keine Repräsentativität beansprucht werden kann. Jedoch wurden die Mikrozensusserhebungen erst 1991 auf die Neuen Bundesländer ausgedehnt, während die Ziehung der Stichprobe für die LV-Ost 71 bereits im Oktober 1990 stattfand. Weil nicht ausgeschlossen ist, dass durch Wanderungen aus und in die Neuen Bundesländer zwischen Oktober 1990 und April 1991 entscheidende Veränderungen in den Randverteilungen hervorgerufen wurden, sind Überlagerungen von Effekten der Stichprobenselektivität und Effekten realer Prozesse möglich. Um dennoch das Ausmaß der Stichprobenverzerrungen der LV-Ost 71 beurteilen zu können, wurden zusätzlich die Mikrozensusserhebungen 1993, 1995 und 1996 zum Vergleich herangezogen. Dabei wurden die Randverteilungen der Mikrozensusserhebungen denen der LV-Ost 71 für die jeweiligen Zeitpunkte (analog zum Mikrozensus Stichmonat: April 1991, April 1993, April 1995, April 1996) gegenübergestellt. Dem liegt die Idee zugrunde, dass durch eine Aneinanderreihung der Querschnittsangaben aus den Mikrozensusserhebungen Prozesse abbildbar sind, die in einer vergleichbaren Form auch in der LV-Ost 71 anzutreffen sein müssten. So sagen zum Beispiel die sich mit der Zeit verändernden Geschlechterproportionen in den Mikrozensusserhebungen etwas über geschlechtsspezifische regionale Mobilität aus, die auch in der LV-Ost 71 sichtbar gemacht werden können. Stichprobenverzerrungen zeigen sich aus dieser Perspektive einerseits in Abweichungen der jährlichen Randverteilungen und andererseits in einem unterschiedlichen Ausmaß der Veränderungen zwischen den jährlichen Randverteilungen.

1.1 Die Lebensverlaufsstudie Ost (Geburtskohorte 1971)

Die Basis für die Erhebungen der Lebensverläufe der 1971er DDR-Geburtskohorte stellte ein Mastersample dar, das im Oktober 1990 aus dem zentralen DDR-Einwohnermelde-register gezogen worden war. Das Mastersample basierte auf einer Gemeindestichprobe, die nach den damals vorhandenen 217 Stadt- und Landkreisen und 10 Gemeindegrößenklassen durch Anordnung geschichtet war. In den Großstädten wurden zusätzlich die Stadtbezirke zur Schichtung herangezogen, so dass sich insgesamt 267 Schichten ergaben. Aus diesem Mastersample wurde eine Personenstichprobe gezogen, die für alle Personen der Grundgesamtheit des angesprochenen Registers die gleiche Auswahlwahrscheinlichkeit hatte. Für die 1971er-Kohorte wurden 2.100 Adressen gezogen, von denen N=1.816 Adressen zum Einsatz gelangten. Mit dem Rückgriff auf dieses Mastersample

konnte sichergestellt werden, dass die regional mobilen Zielpersonen, das heißt jene Personen, die nach dem Oktober 1990 nach Westdeutschland verzogen waren, bei der Ziehung noch berücksichtigt und durch eine Recherche der neuen Adresse auch am neuen Wohnort befragt werden konnten.

Von Mai 1996 bis Januar 1998 konnten 610 Frauen und Männer der 1971er DDR-Geburtskohorte retrospektiv entweder in Telefoninterviews oder in persönlichen Face-to-Face-Interviews auf der Basis eines standardisierten, computergesteuerten Frageprogramms zu ihrem gesamten bisherigen Lebensverlauf befragt werden.⁸ Die Daten wurden während des Interviewgesprächs sofort in Bildschirmmasken eingegeben und in numerischer und alphanumerischer Form (Texte) in einem Rohdatensatz gespeichert. Das Besondere an einer Lebensverlaufserhebung ist, dass die Lebensverläufe im ereignisorientierten Erhebungsdesign erfasst werden, das heißt zum Beispiel zu jeder einzelnen Erwerbstätigkeitsperiode wurden neben einer Reihe anderer Informationen (wie Branche, Anzahl der geleisteten Arbeitsstunden, Einkommen) die Anfangs- und Endzeitpunkte der Tätigkeit auf den Monat genau festgehalten.

Für die Erhebung der Lebensverläufe wurde ein Fragebogen verwendet, der den gesamten Zeitraum von der Geburt der befragten Person bis zum Interviewzeitpunkt erfasste. Der Fragebogen war in sieben Themenbereiche unterteilt, in dem sehr umfangreiche Informationen über den bisherigen Lebensverlauf erkundet wurden. Im ersten Teil (*Herkunftsfamilie*) wurden persönliche Angaben, Daten zur Ausbildung und Erwerbstätigkeit der (Stief-)Eltern sowie Angaben über die Geschwister der befragten Person erfasst. Die *Wohngeschichte* wurde erhoben, in dem alle Wohn- bzw. Aufenthaltsorte der befragten Person in zeitlicher Abfolge erfragt wurden. Für jeden Zeitraum wurden Angaben wie Wohnort, Wohndauer, Wohnart, Größe der Wohnung und im Haushalt lebende Personen erfasst. Hinzu kamen einige Querschnittsangaben zur letzten Wohnung und eine Frage nach dem Zeitpunkt des Auszuges aus dem Elternhaus. Die *Schul- und Berufsausbildung* wurde daran anschließend vollständig, das heißt unter Einschluss sich ergebender Lückenaktivitäten erfasst. Dabei wurde nach der Ausbildungsart, der Dauer der Ausbil-

8 Bei der LV-Ost 71 handelt es sich um eine retrospektive Datenerfassung, das heißt die Befragten mussten ihre bisherige Lebensgeschichte rekonstruieren. Da eine Retrospektivbefragung seitens der Befragten spezifische Fähigkeiten voraussetzt, ist die Qualität von retrospektiven Befragungen in der Umfrageforschung umstritten. Die Befragten müssen in der Lage sein, ihre Vergangenheit chronologisch rekonstruieren zu können und das, was als ihr Lebensverlauf ganzheitlich erlebt (oder zumindest so interpretiert) wird, zeitlich und thematisch zu strukturieren. Dabei ist es naheliegend, einen positiven Zusammenhang zwischen der verstrichenen Zeit und dem Ausmaß von Erinnerungsverzerrungen zu vermuten (vgl. Reimer 2001).

dung und dem Ausbildungsabschluss gefragt. Hinzu kamen noch einige Fragen zum Berufswunsch und dessen Veränderung im Zuge der Wiedervereinigung. Im Abschnitt *Erwerbsgeschichte* wurden für jede berufliche Tätigkeit der befragten Person umfassende Informationen, wie ausgeübte Tätigkeit, Dauer der Beschäftigung, berufliche Stellung, verschiedene Angaben über den Beschäftigungsbetrieb, sowie Arbeitszeit und Einkommen aufgezeichnet. Die Erwerbsgeschichte wurden chronologisch erfasst, das heißt es wurden auch Zeiten der Erwerbslosigkeit (wie Wehr- und Zivildienst, Arbeitslosigkeit) aufgenommen. Darüber hinaus wurden verschiedene Angaben zu Nebentätigkeiten und Aus- und Weiterbildungen erfragt und eine Reihe von Informationen zu Veränderungen während der Wende ermittelt. Daran anschließend wurde die Meinung über verschiedene Aussagen zu Beruf und Erwerb erhoben. Im Abschnitt *Partnerschaften* wurden alle Partner seit dem 16. Lebensjahr erfasst. Dabei wurden neben dem Zeitraum der Partnerschaft auch Angaben zur Ausbildung und Erwerbsgeschichte des Partners während des Zusammenseins ermittelt. Außerdem wurden Informationen zu den Eltern des Partners erhoben und einige Querschnittsfragen zu Partnerschaft und Ehe gestellt. Im darauffolgenden Abschnitt wurden zu jedem – auch nicht-leiblichen – *Kind* der befragten Person neben dem Geburtsdatum und der Ausbildung auch Informationen zu Art und Dauer der Betreuung erhoben. Im Anschluss daran wurden Querschnittsangaben über die Einstellungen zu Familie und Elternschaft ermittelt. Im letzten Teil wurden fast ausschließlich *Querschnittsdaten* für den Zeitpunkt des Interviews bzw. bezogen auf das Jahr 1989 erhoben. Themenschwerpunkte waren Mitgliedschaft in Organisationen und politischen Vereinigungen (retrospektiv erhoben), Einstellungen zu Parteien, Fragen über soziale Netzwerke, das Haushaltseinkommen zum Zeitpunkt des Interviews sowie Zukunftsperspektiven. Die Daten wurden im Anschluss an die Erhebungen in einem mehrstufigem Verfahren einer detaillierten und intensiven Datenprüfung unterzogen (vgl. Matthes/Lichtwardt 2004). Die durch die Datenedition aufgedeckten Widersprüche und Differenzen konnten in einer Nachrecherchephase zum größten Teil geklärt werden (vgl. Lichtwardt 2004).

1.2 Die Mikrozensus-Daten⁹

Der Mikrozensus ist eine amtliche Repräsentativstatistik der Bundesrepublik Deutschland und dient der Bereitstellung statistischer Informationen über die wirtschaftliche und soziale Lage der Bevölkerung sowie über die Erwerbstätigkeit, den Arbeitsmarkt und die

9 Alle Informationen zu den Mikrozensus-Daten sind den Internetseiten des ZUMA, Abteilung Mikrodaten (<http://www.gesis.org/Dauerbeobachtung/Mikrodaten/Daten/Abteilungsdaten/Mikrozensus/mz.htm>) sowie entsprechende Links vom 14. Januar 2003 entnommen.

Ausbildung. Wichtigste Rechtsgrundlage für die Erhebung ist das Gesetz zur Durchführung einer Repräsentativstatistik über die Bevölkerung und den Arbeitsmarkt sowie die Wohnsituation der Haushalte (Mikrozensusgesetz) vom 17. Januar 1996. Die Mikrozensuserhebungen wurden 1991 erstmals auch in den Neuen Bundesländern und Berlin/Ost durchgeführt. Die Grundlage des Mikrozensus bildet eine einstufige, geschichtete 1-Prozent-Haushaltsstichprobe aus bestimmten Auswahlbezirken¹⁰ des Bundesgebietes (laufende Haushaltsstichprobe). Insgesamt nehmen rund 370.000 Haushalte mit 820.000 Personen am Mikrozensus teil; darunter etwa 160.000 Personen in rund 70.000 Haushalten in den Neuen Bundesländern und Berlin/Ost. Im Mikrozensus wird jährlich ein Viertel aller in der Stichprobe enthaltenen Haushalte (beziehungsweise Auswahlbezirke) ausgetauscht, das heißt jeder Haushalt bleibt vier Jahre in der Stichprobe (Verfahren der partiellen Rotation).

Der Mikrozensus wird überwiegend in persönlichen Befragungen durch Interviewer erhoben. Es besteht jedoch auch die Möglichkeit, die Fragen schriftlich zu beantworten. Dabei werden alle in einem Haushalt zusammenlebende Personen befragt beziehungsweise unter Umständen Fremdauskünfte durch andere Haushaltsmitglieder erhoben. Die Angaben – insbesondere zur Erwerbstätigkeit – beziehen sich jeweils auf eine festgelegte Berichtswoche. In der Regel ist dies die letzte feiertagsfreie Woche im April (Berichtswochenkonzept). Der Mikrozensus-Fragebogen umfasst einerseits ein festes Grund- und Ergänzungsprogramm mit jährlich wiederkehrenden Fragestellungen, die überwiegend mit Auskunftspflicht belegt sind. Andererseits gibt es in vierjährigem Rhythmus Zusatzprogramme, die teilweise von der Auskunftspflicht befreit sind. Die Fragen des Grundprogramms beziehen sich unter Anderem auf die Merkmale der Person (wie Alter, Geschlecht, Staatsangehörigkeit), auf den Familien- und Haushaltszusammenhang, auf die Merkmale der Wohnung, auf den Erwerbsstatus (Erwerbstätigkeit, Arbeitslosigkeit, Nichterwerbstätigkeit, Schulbesuch), auf das allgemeine und berufliche Bildungsniveau, auf die Quellen des Lebensunterhalts sowie auf Angaben zur gesetzlichen Renten- und Pflegeversicherung und zur Höhe des Individual- und Haushaltseinkommens. Im jährlichen Ergänzungsprogramm werden auch zusätzliche Fragen zur aktuellen Erwerbstätigkeit gestellt und Angaben zu früheren Erwerbstätigkeitsphasen sowie zur beruflichen und allgemeinen Aus- und Fortbildung erfragt. Im Rahmen der vierjährigen Zusatzprogramme werden unter Anderem Angaben zum Berufs- und Ausbildungspendeln, zur Wohnsituation, zur Krankenversicherung und zur Gesundheit erhoben. Im Mikrozensus

10 Die Definition der westdeutschen Auswahlbezirke beruht auf Auswertungen der Volkszählung 1987. Für die Neuen Bundesländer wurde eine vergleichbare Auswahlgrundlage auf der Basis des „Bevölkerungsregister Statistik“ erstellt. Diese Auswahl wird ständig mit Hilfe der Bautätigkeitsstatistik aktualisiert.

beträgt die Antwortquote in der Regel 97 Prozent. Die Stichprobenausfälle sind überwiegend auf Nichterreichbarkeit zurückzuführen (sogenannte „bekannte Ausfälle“), die im Mikrozensus durch ein besonderes Verfahren (Kompensationsverfahren) ausgeglichen werden.

Das der Forschung zur Verfügung stehende Scientific Use File ist eine 70-Prozent-Stichprobe des Mikrozensus. Um eine Vergleichbarkeit mit den von der amtlichen Statistik veröffentlichten Tabellen zu erreichen, ist eine Gewichtung vorzunehmen (vgl. Schimpl-Neimanns/Müller 2001). Für die Daten 1989-1995 liegt zwar lediglich ein Hochrechnungsfaktor auf Haushaltsebene vor (EF254). Testauswertungen haben jedoch gezeigt, dass die Verwendung des Haushalts-Hochrechnungsfaktors für Auswertungen auf Personenebene nicht zu wesentlich anderen Ergebnissen führt als die Verwendung des korrekten Personen-Hochrechnungsfaktors (Staudenmaier/Granato/Schmidt 2000). Ab 1996 kann neben dem Haushalts-Hochrechnungsfaktor auch auf Hochrechnungsfaktoren für Personen (EF750) zurückgegriffen werden.

1.3 Konzeptionelle Unterschiede zwischen Mikrozensus und Lebensverlaufsstudie und die Konstruktion der Vergleichspopulation

Die Zielpopulation der LV-Ost 71 besteht aus Personen der Geburtskohorte 1971, die im Oktober 1990 im zentralen Einwohnermelderegister der ehemaligen DDR verzeichnet waren und bei der Befragung 1996/97 in Ost- oder Westdeutschland lebten (insgesamt N=610 Personen). In der retrospektiven Längsschnitterhebung LV-Ost 71 stehen Informationen über die befragten Personen von der Geburt bis zum Interviewzeitpunkt, also mindestens bis Mai 1996, zur Verfügung. Die Vergleichspopulation aus dem Mikrozensus sollte möglichst ähnlich ausgewählt werden.

Bei der Konstruktion der Vergleichspopulation aus den Mikrozensus-erhebungen waren folgende konzeptionellen Unterschiede zwischen Mikrozensus und LV-Ost 71 zu beachten:

- *Unterschiedliche Zielpopulation:* Es war kein Problem, im Mikrozensus die deutschen Staatsangehörigen des Geburtsjahrgangs 1971 aus den privaten Haushalten zu identifizieren. Eine Einschränkung auf „Bürger der DDR im Jahre 1990“ war allerdings nicht vergleichbar möglich. Für den Vergleich der LV-Ost 71 mit dem Mikrozensus wurden daher alle Personen herangezogen, die zur jeweiligen Mikrozensus-Befragung in den Neuen Bundesländern (bzw. den Ostbezirken Berlins) lebten. Regional gesehen bedeutet dies eine Einschränkung auf die Bevölkerung der Neuen Bundesländer, so dass einerseits in den Neuen Bundesländern lebende Westdeutsche in der Auswahlpopulation des Mikrozensus auftauchen können und andererseits die in die Alten

Bundesländer umgezogenen Ostdeutschen in der Auswahlpopulation des Mikrozensus nicht enthalten sind. Vereinfacht gesagt, beinhaltet die LV-Ost 71 Ost-West- und keine West-Ost-Wanderer. Bei den Mikrozensus-Vergleichsdateien ist dies genau umgekehrt.¹¹

- *Unterschiedliche Stichprobeneinheiten:* Der Mikrozensus ist eine Haushaltsstichprobe, bei der alle in einem Haushalt lebenden Personen entweder direkt befragt werden bzw. eine erwachsene Person über die weiteren (alle minderjährigen) Haushaltsmitglieder Auskunft gibt. Im Gegensatz dazu beruht die LV-Ost 71 auf einer Personenstichprobe, die wiederum auf einer Gemeindestichprobe basiert. Inwiefern sich durch diese Unterschiede bei der Stichprobenziehung verschiedene Auswahlwahrscheinlichkeiten für die 1971 in der DDR Geborenen ergeben haben, muss anderen Untersuchungen vorbehalten bleiben. Bei der Ermittlung der Randverteilungen des Mikrozensus sollten jedoch die angegebenen Hochrechnungsfaktoren einbezogen werden.
- *Unterschiedliches Methodendesign:* Dadurch, dass die LV-Ost 71 eine retrospektive Längsschnitterhebung ist, sind die Angaben für jeden Befragten für jeden Zeitpunkt rekonstruierbar. Dagegen ist der Mikrozensus eine wiederholte repräsentative Querschnittsbefragung, in dem die Befragten – sofern sie nicht umziehen – jeweils 4 Jahre verbleiben und jedes Jahr $\frac{1}{4}$ der Haushalte ausgetauscht wird (rotierendes Panel). Mit der LV-Datei kann eine konstante Population beobachtet werden, während die Zielpopulation im Mikrozensus ständig wechselt. Allerdings kann die LV-Ost 71 nur für den Zeitpunkt der Stichprobenziehung (also Oktober 1990) Anspruch auf Repräsentativität erheben, während der Mikrozensus zum jeweiligen Erhebungszeitpunkt (also April 1991, April 1993, April 1995, April 1996) repräsentativ ist.
- *Unterschiedliche Variablen:* In der Regel konnten Variablen mit ihren Ausprägungen aus dem Mikrozensus nicht direkt mit Variablen aus der LV-Ost 71 verglichen werden. Bei nahezu allen Variablen besteht nicht nur ein Unterschied zwischen den Kategorien

11 Für die sozialwissenschaftliche Forschung wäre es wichtig, wenn im Mikrozensus auch Informationen zur regionale Herkunft der Personen verfügbar wären. Nicht nur um bei Vergleichen mit dem Mikrozensus eine bessere Übereinstimmung der Zielpopulationen zu erreichen, sondern auch um Ost-West-Wanderungen oder andere regionale Wanderungsbewegungen gezielter untersuchen zu können. Der Fragenkomplex, der sich auf einen im Vergleich zum vorangegangenen Jahr vorgenommenen Wohnsitzwechsel bezieht, ist durch Rotieren des Panels für eine eindeutige Identifikation von Ost-West-Wanderern nicht ausreichend.

der LV-Ost 71 und dem Mikrozensus, auch im Mikrozensus selbst werden häufig Variablen und deren Kategorien modifiziert.

- *Weitere Unterschiede (Erhebungsmethode, Teilnahme):* Im Gegensatz zu wissenschaftlichen Untersuchungen wie der LV-Ost 71, an der die befragten Personen freiwillig teilnehmen, besteht für die meisten Fragen im Mikrozensus gesetzliche Auskunftspflicht. Bei einer Response-Rate von etwa 97 Prozent ist im Mikrozensus mit einer höheren Repräsentativität zu rechnen. Das Erhebungsverfahren des Mikrozensus ist überwiegend ein mündliches Interview und teilweise ein schriftlich selbst ausgefüllter Fragebogen (1996: 80 Prozent mündlich; 20 Prozent Selbstausfüller). Hingegen wurden in der LV-Ost 71 CATI- und CAPI¹²-Interviewprogramme verwendet, die Filterführungen und Konsistenzprüfungen der Daten bereits bei der Erhebung ermöglichten.

In der folgenden Tabelle sind die oben genannten Unterschiede zwischen den Studien noch einmal zusammen gefasst.

Tabelle 1: Konzeptionelle Unterschiede zwischen Mikrozensus und Lebensverlaufsstudie

	Mikrozensus	LV-Ost 71
Ausgewählte Zielpopulation	Geburtsjahrgang 1971 deutsche Personen in Privathaushalten, die im Erhebungsjahr in den Neuen Bundesländern lebten	Geburtsjahrgang 1971 ostdeutsche Personen in Privathaushalten, die im Oktober 1990 Einwohner der DDR waren
Stichprobeneinheiten	Haushalte	Personen
Methodendesign	wiederholte Querschnittsbefragung; 70-Prozent-Substichprobe als Public Use File einer 1-Prozent-Bevölkerungsstichprobe	retrospektive Längsschnittbefragung; durchgeführt 1996/97
Variablen	Variablen und Kategorien nicht identisch mit LV-Ost 71 und im Zeitverlauf häufig modifiziert	Variablen und Kategorien nicht identisch mit Mikrozensus
Erhebungsmethode/	überwiegend mündliches Interview, teilweise schriftliche Selbstausfüller	Computerunterstütztes Befragungsprogramm über Telefon (CATI) oder als persönliches Interview (CAPI)
Teilnahme	überwiegend Auskunftspflicht	freiwillige Teilnahme

12 CATI=computer assisted telephone interview; CAPI=computer assisted personal interview

2. Vergleich der LV- Ost 71 mit den Erhebungen des Mikrozensus

Aus dem Mikrozensus wurden für den Vergleich mit der LV-Ost 71 die Daten aller 1971 geborenen, zum jeweiligen Zeitpunkt in den neuen Bundesländern (bzw. in den Ostbezirken Berlins) lebenden Personen mit deutscher Staatsangehörigkeit, die in privaten Haushalten lebten, herangezogen (vgl. Tabelle 2).

Tabelle 2: Definition der Auswahlpopulation¹³

	MZ91	MZ93	MZ95	MZ96	LV-Ost 71
Geburtskohorte	EF37=71	EF37=71	EF37=71	EF33=1971	1971
Ost-Berlin	EF1=11 & EF8=9	EF1=11 & EF8=9	EF1=11 & EF8=9	EF1=11 & EF708=9	
Brandenburg	EF1=12	EF1=12	EF1=12	EF1=12	
Mecklenburg	EF1=13	EF1=13	EF1=13	EF1=13	
Sachsen	EF1=14	EF1=14	EF1=14	EF1=14	
Sachsen-A.	EF1=15	EF1=15	EF1=15	EF1=15	
Thüringen	EF1=16	EF1=16	EF1=16	EF1=16	
Deutsche Staatsangehörigkeit	EF41=1	EF41=1	EF41=1	EF43=1	alle
Personen in privaten Haushalten	EF27=1	EF27=1	EF27=1	EF506=1	alle
N ungewichtet	1.391	1.182	1.205	1.164	610
N gewichtet	2.177	1.842	1.898	1.881	

Beim Vergleich der LV-Ost 71 mit dem Mikrozensus 1991, 1993, 1995 und 1996 wurden neben demographischen Grundinformationen vor allem die Randverteilungen solcher Variablen untersucht, die für die Lebensverlaufsstudien wichtige Themen zur Familiengründung und zu den Bildungs- und Erwerbsverläufen betreffen. Allerdings konnten nur solche Variablen verglichen werden, die in beiden Studien hinreichend kompatibel waren. Der Ausgangspunkt für die Auswahl der Variablen stellte der Mikrozensus dar. Wenn die

13 Dabei steht im Folgenden MZ91 für Mikrozensus 1991, MZ93 für Mikrozensus 1993, MZ95 für Mikrozensus 1995, MZ96 für Mikrozensus 1996, LV91 für den erzeugten Querschnitt des Jahres 1991 aus der LV-Ost 71, LV93 für den des Jahres 1993, LV95 für den des Jahres 1995 und LV96 für den des Jahres 1996.

Variablen inhaltlich vergleichbar waren, wurden die Variablen der LV-Ost 71 entsprechend dem Zuschnitt der Variablen im Mikrozensus konstruiert (vgl. Tabelle 3).

Tabelle 3: Übersicht über die verwendeten Variablen

Verwendete Variablen		Variablenbezeichnung in		
		MZ91; MZ93; MZ95	MZ96	LV-Ost 71
Demographische Basisvariablen	Geschlecht	EF35	EF32	sex
	Bundesland	EF1	EF1	bl91, bl93, bl95, bl96
	Gemeindegrößenklasse (MZ) bzw. Einwohnerzahl (LV)	EF8	EF708	ez91, ez93, ez95, ez96
Haushalt und Familie	Familienstand	EF38	EF35	fs91, fs93, fs95, fs96
	Eheschließungsjahr	EF39	EF36	ej91, ej93, ej95, ej96
	Kinder	EF173 EF174EF175	EF528 EF529 EF530	k91, k93, k95, k96
	Stellung der Person im Haushalt	EF30	EF509	sh91, sh93, sh95, sh96
Bildung und Ausbildung	Höchster Schulabschluss	EF121	EF287	sa91, sa93, sa95, sa96
	Höchster Ausbildungsabschluss		EF289 EF290	hab
Beruf und Erwerb	Erwerbsbeteiligung	EF34=1	EF338U2=1	et91, et93, et95, et96
	Arbeitslosigkeit	EF34=5	EF338U2=2	al91, al93, al95, al96
	Berufliche Stellung	EF110	EF167	bs91, bs93, bs95, bs96
Einkommen	Haushaltsnettoeinkommen		EF539	Eink

Da man davon ausgehen kann, dass die wichtigsten sozialen Merkmale regional mobiler Personen sehr ähnlich sind und diese im Mikrozensus nicht identifizierbar sind, wurden die nach Westdeutschland (inklusive Berlin/West) umgezogenen bzw. pendelnden Perso-

nen¹⁴ nur in Zweifelsfällen aus der LV-Ost 71 ausgeschlossen. Bei den Analysen der Mikrozensusdaten wurde im Folgenden immer die jeweiligen Hochrechnungsfaktoren verwendet.

Für die vergleichbaren Variablen wurden Querschnittsverteilungen für die gleichen Zeitpunkte ermittelt, die für den Mikrozensus zur Verfügung standen. Trotz inhaltlich gleicher Bedeutung der Variable gab es dennoch mitunter unterschiedliche Kodierungen, die nachträglich nicht in ein vergleichbares Kodierschema übertragen werden konnten. In solchen Fällen wurden die Randverteilungen für die jeweiligen unterschiedlichen Kodierungen belassen.

2.1 Vergleich univariater Verteilungen

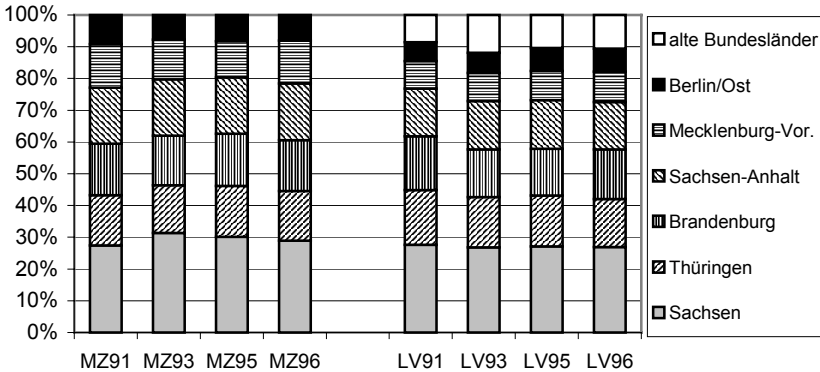
2.1.1 Demographische Basisvariablen

Vergleicht man die Anteile von Männern und Frauen in den einzelnen Studien zeigen sich keine Unterschiede zwischen den Mikrozensusserhebungen und der LV-Ost 71. Allerdings weist die LV-Ost 71 im April 1991, 1993 und 1995 einen leicht höheren und im April 1996 einen leicht niedrigeren Anteil an Männern auf als die Erhebungen des Mikrozensus. Anhand der Chi-Quadrat-Werte aus dem Anpassungstest ist jedoch ersichtlich, dass diese Unterschiede nicht signifikant sind. Das Verhältnis von Männern und Frauen ist also in der LV-Ost 71 in sehr guter Näherung erhoben worden.

Die unterschiedliche Zielpopulation des Mikrozensus und der LV-Ost 71 wird besonders augenfällig, wenn man die prozentualen Anteile der Bundesländer, in denen die 1971 Geborenen jeweils wohnten, betrachtet (vgl. Abbildung 1).

14 Diese spezifische Personengruppe wird im folgenden Ost-West-Wanderer genannt.

Abbildung 1: Wohnort – Bundesländeranteile



Chi-Quadrat: 1991=329,8***, 1993=299,7***; 1995=265,1***; 1996=271,8***

Vor diesem Hintergrund überrascht es nicht, dass der Chi-Quadrat-Test die Abweichungen zwischen Mikrozensus und LV-Ost 71 als signifikant ausweist.¹⁵ Stellt man dem Mikrozensusserhebungen dagegen nur die in den Neuen Bundesländern Wohnenden der LV-Ost 71 gegenüber, sind die Relationen der Bundesländer sehr ähnlich. Abgesehen von einem geringeren Anteil der in Mecklenburg-Vorpommern Wohnenden im Jahre 1991, sind keine signifikanten Unterschiede zwischen Mikrozensus und LV-Ost 71 zu beobachten.

Die Variable „Gemeindegroßenklasse“ lag im Mikrozensus in Kategorien vor, die in den vier Vergleichsjahren unterschiedlich gebildet waren. Deshalb musste eine sehr grobe Kategorisierung („unter 20.000 Einwohner“, „20.000 bis unter 500.000 Einwohner“, „500.000 Einwohner und mehr“) vorgenommen werden. Dementsprechend wurde die Kategorisierung der Angaben der Befragten über die Einwohnerzahl der jeweiligen Wohnorte aus der LV-Ost 71 gebildet. Die Kategorie „unter 20.000 Einwohner“ weist für alle vier Beobachtungszeitpunkte eine hohe Übereinstimmung zwischen Mikrozensus und LV-Ost 71 aus. Dass die Abweichungen zwischen Mikrozensus und LV-Ost 71 trotzdem in

15 Dabei nahm der Anteil der in Westdeutschland lebenden bzw. pendelnden Befragten in der LV-Ost 71 bis 1993 zu. 1995 ist ein leichter Überhang der Rückwandererquote und danach eine stabile Wanderungsquote festzustellen. Allerdings kann mit dieser Darstellung nichts darüber ausgesagt werden, ob es sich bei den in Westdeutschland lebenden Ostdeutschen in allen Vergleichsjahren um die gleiche Personengruppe oder um einen wechselnden Personenkreis handelt, der von Ost nach West und wieder zurück gezogen war.

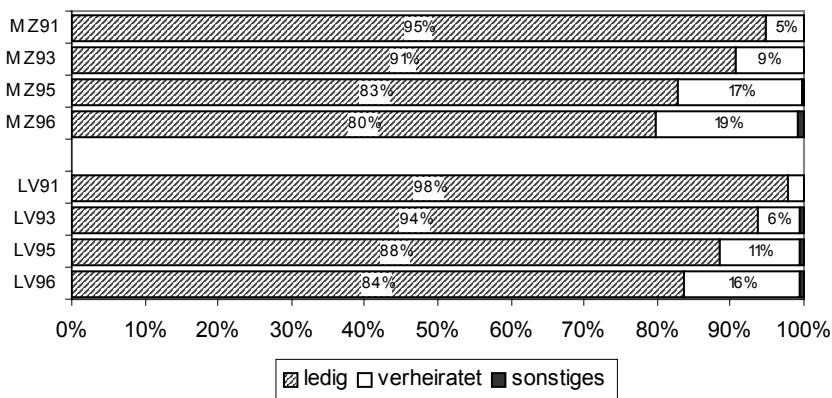
den Jahren 1991 und 1992 signifikant sind, ist vordergründig auf den höheren Anteil der Kategorie "über 500.000 Einwohner" im Mikrozensus 1991 und 1993 zurückzuführen. Warum aber sinkt im Mikrozensus dieser Anteil zwischen 1993 und 1995 so deutlich? Recherchen haben ergeben, dass die ostdeutschen Großstädte in dieser Zeit eine große Anzahl ihrer Einwohner verloren haben und zum Beispiel Leipzig nicht mehr zu den Städten mit über 500.000 Einwohner zählt. Die signifikanten Differenzen im Jahre 1991 und 1993 sind deshalb wohl eher auf Schwierigkeiten der Befragten in der LV-Ost 71 bei der retrospektiven Einschätzung der Wohnortgröße zurückzuführen.

2.1.2 Haushalt und Familie

Um die Abweichungen zwischen den beiden Erhebungen hinsichtlich der Haushaltsstruktur zu analysieren, wurde die Stellung der befragten Person im Haushalt verglichen. Dabei wurden im Mikrozensus alle Haushalte der Befragten, die sich entweder als Haushaltsvorstand oder Ehefrau/-mann des Haushaltsvorstandes bezeichneten, als eigene Haushalte verstanden. Dem wurde die Angabe der Befragten, ob es sich bei der jeweiligen Wohnperiode um einen eigenen Haushalt gehandelt hatte, gegenübergestellt. Der so definierte Anteil der Personen mit eigenem Haushalt weist für alle Vergleichsjahre eine sehr hohe Übereinstimmung zwischen beiden Erhebungen auf. Etwa 27 Prozent der Befragten hatten in beiden Studien bereits 1991 einen eigenen Haushalt, 1996 waren es etwa 69 Prozent.

Es lassen sich auch bei einer Differenzierung nach Geschlecht bzw. nach Bildungsniveau keine interpretationswürdigen Unterschiede feststellen. Die Vermutung, dass die Ost-West-Wanderer die Anteile der Personen im eigenen Haushalt „hochtreiben“ würden, bestätigt sich ebenfalls nicht. Vielmehr erhöht sich der Anteil der Personen im eigenen Haushalt in der LV-Ost 71 bei Nichtberücksichtigung der Ost-West-Wanderer. Das Ausmaß dieser Steigerung verringert sich jedoch mit über die Jahre abnehmendem Gewicht. Offensichtlich verzögerte sich bei den Ost-West-Wanderern durch ihre regionale Mobilität auch die Gründung eines eigenen Haushaltes.

In allen vier Vergleichsjahren weist die LV-Ost 71 einen höheren Anteil von ledigen Personen auf als im Mikrozensus. Von den in der LV-Ost 71 befragten Personen waren im April 1996 noch 83,7 Prozent ledig, im Mikrozensus nur 79,8 Prozent. In Abbildung 2 wird die Verteilung nach Familienstand für beide Studien getrennt ausgewiesen, weil dadurch gezeigt werden kann, dass in der LV-Ost 71 zwar insgesamt weniger Verheiratete vertreten sind, aber die Abnahme der Anzahl der Ledigen in beiden Studien fast synchron erfolgte.

Abbildung 2: Familienstand¹⁶

Chi-Quadrat: 1991=10,6***; 1993=16,3***; 1995=22,9***; 1996=4,7*

Sowohl der Anteil lediger Frauen als auch der Anteil lediger Männer ist in der LV-Ost 71 in allen vier Vergleichsjahren höher als im Mikrozensus. Die prozentualen Anteile lediger Personen liegen in der LV-Ost 71 bei den Männern zwischen 1,5 und 3,8 Prozentpunkten, bei den Frauen zwischen 2,8 und 7,9 Prozentpunkten über den Mikrozensuswerten. Übereinstimmend zeigt sich in beiden Studien ein allmählicher Trend zur Heirat, der bei den Frauen deutlich stärker ausgeprägt ist.

Die Vermutung, dass diese Stichprobenverzerrungen vordergründig auf Wanderungen junger, lediger Ostdeutscher zurückgeführt werden könnten, lässt sich nicht bestätigen. Die Abweichungen zwischen Mikrozensus und LV-Ost 71 hinsichtlich des Familienstandes bleiben auch dann meist signifikant, wenn die Ost-West-Wanderer aus der Betrachtung ausgeschlossen werden. Bei den Ost-West-Wanderern handelt es sich also nicht häufiger um Ledige.

Die Analyse des Eheschließungsjahres liefert einen ersten Anhaltspunkt dafür, dass sich die vermeintliche Stichprobenverzerrung nach Familienstand als Bildungseffekt entpuppen könnte. Der höhere Anteil an verheirateten Personen im Mikrozensus ist auf eine Unterrepräsentation sehr früher Eheschließungen in der LV-Ost 71 zurückzuführen. So ergab sich zum Beispiel für das Eheschließungsjahr 1989 – die 1971 Geborenen waren zu

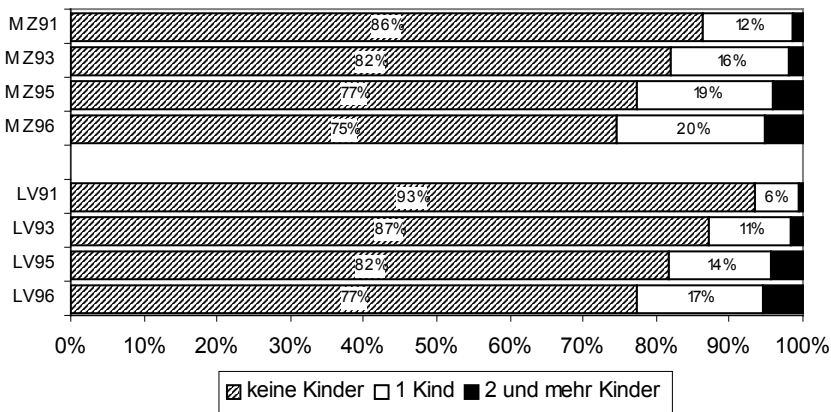
16 Die Kategorie "sonstiges" umfasst Verwitwete oder Geschiedene.

diesem Zeitpunkt gerade 18 Jahre alt – in allen vier Mikrozensus-Erhebungen ein Anteil zwischen 0,7 und 1,3 Prozent verheirateter Personen, in der LV-Ost 71 dagegen von nur 0,5 Prozent. Für das Eheschließungsjahr 1990 waren im Mikrozensus 1995 bereits 2,7 Prozent, im Mikrozensus 1996 2,2 Prozent der Personen verheiratet. In der LV-Ost 71 waren es im selben Jahr nur 1,7 Prozent. Davon ausgehend, dass höher Gebildete tendenziell später heiraten, würden daraus auch dementsprechende Abweichungen zwischen Mikrozensus und LV-Ost 71 hinsichtlich des Familienstandes resultieren. Diese Vermutung muss sich allerdings erst noch in einer multivariaten Analyse bestätigen.

Die Variablen, die über die Kinder der Befragten Auskunft geben, sind zwischen den beiden Vergleichsstudien nur eingeschränkt vergleichbar. Weil im Mikrozensus die verwandtschaftliche Beziehung zur Haushaltsbezugsperson, jedoch nicht zwischen den einzelnen Haushaltsmitgliedern erhoben werden, entstehen vor allem bei Mehrgenerationen-Haushalten Probleme bei der Identifikation der Kinder der Befragten. Lebt zum Beispiel eine junge Frau mit ihrem Kind bei ihrer Schwester, könnte man ihr das Kind bei Verwendung der Individualinformationen nicht zuordnen. Benutzt man dagegen die Bandsatzerweiterungen (Anzahl der Kinder im Haushalt) könnten die Kinder der Geschwister oder die Kinder anderer Haushaltsmitglieder als eigene Kinder betrachtet werden. Demgegenüber ist es mit den Daten der LV-Ost 71 möglich, eindeutig zwischen leiblichen Kindern, Kindern des Partners und Geschwistern zu unterscheiden sowie die genaue Haushaltszusammensetzung aus der Wohngeschichte zu rekonstruieren.

Da die 1971 Geborenen 1996 erst 25 Jahre alt waren und wahrscheinlich nur wenige Geschwister unter 10 Jahren hatten, wurden für den folgenden Vergleich die Anzahl der Kinder im Haushalt unter 10 Jahren herangezogen. Aus der LV-Ost 71 wurden nur jene Kinder in den Vergleich einbezogen, die in der auch in der Wohngeschichte genannt worden waren.¹⁷ Legt man dem Vergleich diese Zuordnungskriterien zugrunde, ist die Anzahl der Kinder im Mikrozensus in allen Vergleichsjahren höher als in der LV-Ost 71 (vgl. Abbildung 3).

17 Diese Einschränkung betraf fast ausschließlich Männer, die zwar eigene Kinder hatten, mit diesen jedoch nicht in einem Haushalt zusammen lebten.

Abbildung 3: Kinderanzahl

Chi-Quadrat: 1991=23,2***; 1993=9,9**; 1995=8,8*; 1996=5,3

Bei den Männern fielen die Unterschiede hinsichtlich der Kinderanzahl zwischen Mikrozensus und LV-Ost 71 im Jahre 1991 zwar noch recht hoch aus (MZ91: 88,6 Prozent, LV91: 96,6 Prozent), sind danach jedoch nicht mehr signifikant. Dagegen bleiben die Abweichungen bei den Frauen fast stabil und sanken zu keinem Vergleichszeitpunkt unter 6 Prozent. Demzufolge sind Frauen mit Kindern und Männer, die früh Väter wurden, in der LV-Ost 71 unterrepräsentiert. Dabei hatten sowohl die Männer als auch die Frauen, die in die Alten Bundesländer umgezogen sind bzw. sich dort überwiegend aufhielten, seltener Kinder als die in den Neuen Bundesländern Wohnenden. Jedoch werden selbst bei Nichtberücksichtigung der Ost-West-Wanderer in der LV-Ost 71 diese Unterschiede nicht vernachlässigbar klein.

Äußerst bemerkenswert ist dabei, dass der Anteil der Kinderlosen – ähnlich wie der Ledigenanteil beim Familienstand – fast synchron abnimmt. Möglicherweise ist dies ein Indiz dafür, dass beide Verzerrungsbefunde auf den gleichen Effekt zurückzuführen sind, nämlich darauf, dass Niedrigqualifizierte seltener an freiwilligen sozialwissenschaftlichen Untersuchungen teilnehmen als Höherqualifizierte. Diese Vermutung lässt sich jedoch erst durch eine multivariate Analyse überprüfen.

2.1.3 Bildung und Ausbildung

Die Zuordnung der in der LV-Ost 71 erhobenen Schul- und Ausbildungsabschlüsse zu den entsprechenden Kategorien im Mikrozensus waren nicht immer zweifelsfrei möglich. Die Frage nach dem höchsten allgemeinbildenden Schulabschluss konnten die Befragten im

Mikrozensus beantworten, indem sie eine der vorgegebenen Kategorien ankreuzten. Dabei standen neben dem Abschluss einer Polytechnischen Oberschule (POS) (leider nicht differenziert nach 8. Klasse- oder 10. Klasse-Abschluss) und dem Abitur, wobei leider nicht nach Abschluss der Erweiterten Oberschule (EOS) oder der Berufsausbildung mit Abitur (BmA) differenziert wurde, auch ein Hauptschulabschluss oder der Realschulabschluss zur Auswahl. Demgegenüber wurde in der LV-Ost 71 die gesamte schulische Laufbahn – einschließlich der Schulbesuche nach dem Verlassen der allgemeinbildenden Schule – erhoben.

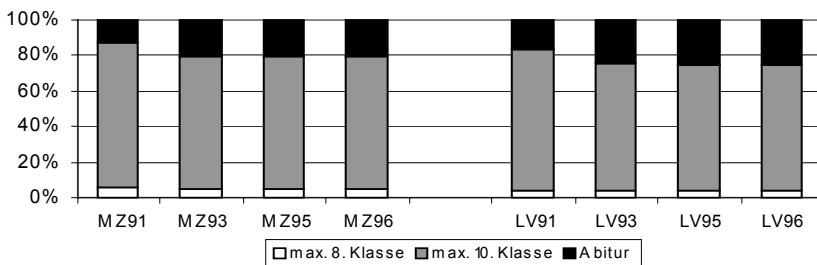
Weil die Zuordnung der in der LV-Ost 71 erhobenen Schulabschlüsse zu den entsprechenden Kategorien im Mikrozensus nicht zweifelsfrei möglich war, wurden die Verteilungen der Schul- und Ausbildungsabschlüsse zunächst nur gegenüber gestellt. Dabei ist der relativ hohe Anteil von Personen mit „westdeutschen“ Schulabschlüssen im Mikrozensus besonders erstaunlich, denn die 1971 Geborenen konnten vor 1989 keine Haupt- oder Realschulabschlüsse erworben haben, und es war in der Regel auch nicht notwendig, einen Hauptschul- oder Realschulabschluss nachzuholen, da fast alle DDR-Schulabschlüsse nach 1989 anerkannt wurden. Vor diesem Hintergrund kann dieser Befund zweierlei bedeuten. Entweder ordneten die Befragten des Mikrozensus ihre erreichten Schulabschlüsse in die westdeutschen Kategorienmuster ein. Oder es zog tatsächlich ein dementsprechend großer Anteil junger Westdeutscher nach Ostdeutschland um, so dass die Kategorien aufgrund der spezifischen Auswahlpopulation des Mikrozensus so stark vertreten war.¹⁸

Vergleicht man dennoch die höchsten allgemeinbildenden Schulabschlüsse, indem alle im Mikrozensus angegebenen Realschulabschlüsse als 10.Klasse-Schulabschlüsse gewertet werden und jene Personen aus dem Mikrozensus, die angaben das Abitur, das Fachabitur oder die Fachhochschulreife vorweisen zu können sowie jene Personen aus der LV-Ost 71, die das Abitur entweder auf einer EOS, während einer Berufsausbildung mit Abitur erworben oder das Abitur später nachgeholt haben, als Abiturienten betrachtet werden, wird deutlich, dass Personen mit niedrigeren Schulabschlüssen in der LV-Ost 71 unter- und Personen mit höheren Schulabschlüssen überrepräsentiert sind. In allen Vergleichsjahren hatte in der LV-Ost 71 ein geringerer Anteil der Befragten maximal einen 8.Klasse-Schulabschluss als im Mikrozensus (LV-Ost 71: 4 Prozent; Mikrozensus: 5-6 Prozent). Vergleicht man die Abiturientenanteile wird deutlich, dass Abiturienten in der LV-Ost 71

18 Darüber hinaus haben, obwohl die Frage nach dem höchsten Schulabschluss im Mikrozensus ab 1996 für Personen unter 51 Jahren nicht mehr freiwillig ist, immerhin 3,2 Prozent der 1971 Geborenen, die 1996 in Ostdeutschland lebten, keine Angabe dazu gemacht. Unverständlich ist uns, warum (auch über 15-jährige) Personen, die sich zum Befragungszeitpunkt in einer allgemeinbildenden Ausbildung befinden, nicht nach ihrem bis dahin erworbenen Schulabschluss gefragt werden und warum es einem beachtlichen Teil der Befragten nicht auffällt, dass die Beantwortung dieser Frage für Unter-51-Jährige nicht freiwillig ist.

mit einem Anteil von fast 25 Prozent im Vergleich zum Mikrozensus mit etwa 20 Prozent deutlich überrepräsentiert sind (vgl. Abbildung 4).¹⁹

Abbildung 4: Höchster allgemeinbildender Schulabschluss



Chi-Quadrat: 1991=9,0***; 1993=5,7**, 1995=9,1***; 1996=9,9***

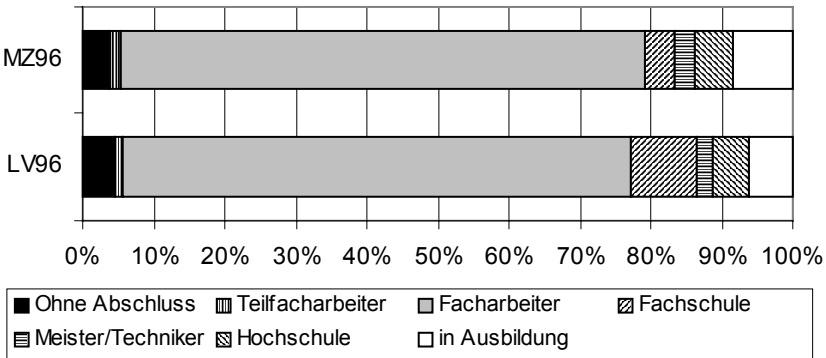
Im Mikrozensus hatten in allen Vergleichsjahren bedeutend mehr Frauen als Männer das Abitur. Zwar lag der Anteil der Frauen mit Abitur im Jahr 1991 noch 3 Prozent über dem Anteil der Männer mit Abitur, aber schon 1993 kehrten sich die Relationen um. 1996 hatten in der LV-Ost 71 27 Prozent der Männer und 25 Prozent der Frauen das Abitur. Dies ist vordergründig darauf zurückzuführen, dass Männer häufiger als Frauen eine 3-jährige Berufsausbildung mit Abitur absolvierten, wohingegen die Frauen häufiger nach einem 2-jährigen Besuch der EOS das Abitur ablegten. An der LV-Ost 71 nahmen männliche Abiturienten überproportional häufig teil.

Der Vergleich der höchsten beruflichen Ausbildungsabschlüsse war nur für das Jahr 1996 möglich, da im Mikrozensus in den Jahren zuvor nicht nach dem höchsten, sondern nach dem letzten beruflichen Ausbildungsabschluss gefragt wurde. Aufgrund der unterschiedlichen Kategorisierung des beruflichen Ausbildungsabschlusses in Mikrozensus und LV-Ost 71 werden Personen mit „Anlernausbildung“ im Mikrozensus den „Teilfacharbeitern“ in der LV-Ost 71, Personen mit „Lehre“ im Mikrozensus den „Facharbeitern“ in der LV-Ost 71, Personen mit einem Abschluss als „Meister/Techniker“ oder „Fachschule (ehem. DDR)“ im Mikrozensus den „Meistern“ und „Technikern“ in der LV-Ost 71 sowie

¹⁹ Dagegen wird in der amtlichen Statistik ausgewiesen, dass 1989 in der DDR 10,3 Prozent der Schulabgänger die Schule mit einem niedrigeren Abschluss als dem POS-10.Klasse-Abschluss verlassen hatten (Bundesministerium für Bildung und Wissenschaft 1990/91: 363). Im Mikrozensus wird dieser Anteil vermutlich aufgrund der Zuordnungsschwierigkeiten der Realschulabschlüsse unterschätzt. Der Abiturientenanteil wird in dieser Statistik mit 11,4 Prozent beziffert (ebd.).

die (Fach-) Hochschulabsolventen beider Erhebungen gegenübergestellt (vgl. Abbildung 5).²⁰

Abbildung 5: Höchster Ausbildungsabschluss 1996²¹



Chi-Quadrat: 15,0**

Das Überraschendste an dieser Abbildung ist, dass der Anteil der Befragten ohne Ausbildungsabschluss und mit einer Anlernausbildung sowie der Anteil der (Fach-) Hochschulabsolventen in beiden Erhebungen nahezu identisch ist. Beim Vergleich der Schulabschlüsse hingegen war festgestellt worden, dass an der LV-Ost 71 Personen mit Abitur überproportional teilgenommen hatten. Diese Übereinstimmung kann entweder dadurch zustande kommen, dass viele Abiturienten zum Vergleichszeitpunkt noch studieren, oder weil Befragte, die bislang noch ohne Ausbildungsabschluss geblieben sind, eine Ausbildung absolvieren. Deshalb kann damit nicht eindeutig belegt werden, ob in der LV-Ost 71 Personen ohne beruflichen Ausbildungsabschluss unterrepräsentiert sind.

Die große Differenz bei den Fachschulabsolventen ist auch nicht einwandfrei aufzuklären. Unklar ist vor allem, ob die Differenz auf Unklarheiten bei der Zuordnung der absolvierten Ausbildung durch die Zielpersonen oder auf tatsächliche Abweichungen zurückzuführen ist. In der DDR konnten viele Berufe (vor allem im medizinischen,

20 Um die noch in Ausbildung befindlichen Befragten im Mikrozensus zu identifizieren, wurde die Angabe zum gegenwärtigen Schul- beziehungsweise Hochschulbesuch (EF71 und EF295) verwendet.

21 Für den Vergleich mussten 3,6 Prozent der 1971 geborenen Befragten des Mikrozensus aus der Betrachtung ausgeschlossen werden, weil für sie keine Angabe zum höchsten beruflichen Ausbildungsabschluss vorlag, obwohl die Frage nach dem höchsten Ausbildungsabschluss für Personen unter 51 Jahren ab 1996 nicht mehr freiwillig ist.

pädagogischen und künstlerischem Bereich) nur in Fachschulen erlernt werden. Eine Reihe dieser Fachschulausbildungen konnten nach 1989 jedoch auch im Rahmen einer dualen Ausbildung absolviert werden. Es ist denkbar, dass der höhere Anteil von Personen mit einer „Lehre“ im Mikrozensus mit dem niedrigeren Anteil von Fachschulabsolventen im Zusammenhang steht. Darüber hinaus sind einige DDR-Fachschulabschlüsse nicht anerkannt worden (zum Beispiel Pionierleiter), die in der LV-Ost 71 dennoch als Fachschulabschlüsse gewertet, im Mikrozensus jedoch von den Befragten nicht unbedingt als berufliche Ausbildungsabschlüsse angegeben wurden.

Da Fachschulausbildungen vorwiegend von Frauen absolviert wurden, ist es angesichts der vermuteten Zuordnungsschwierigkeiten nicht verwunderlich, dass bei den Frauen signifikante Unterschiede hinsichtlich der höchsten Ausbildungsabschlüsse zu finden sind. Der höhere Fachschulabsolventenanteil bei den Frauen in der LV-Ost 71 korrespondiert mit einem höheren Anteil von Lehrabsolventinnen, Meisterinnen/Technikerinnen sowie Frauen ohne Ausbildungsabschluss. Demgegenüber sind die Unterschiede bei den Männern nicht signifikant. An dieser Stelle kann schon festgestellt werden, dass die größeren Abweichungen zwischen LV-Ost 71 und Mikrozensus bei den Frauen wohl auch als ein Hinweis auf größere Verwerfungen in den weiblichen Lebensverläufen während der ostdeutschen Transformation zu verstehen sind.

2.1.4 Erwerbs- und Berufsstatus

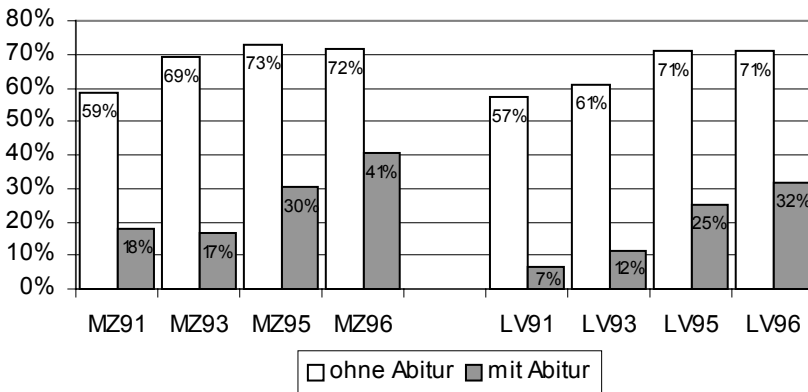
Der Vergleich der Variablen zu Erwerbstätigkeit und Arbeitslosigkeit sah sich mit besonderen Definitionsproblemen der Begriffe Erwerbstätigkeit und Arbeitslosigkeit konfrontiert. Während der Mikrozensus alle Personen, die im Berichtsmonat einer Beschäftigung von mindestens einer Stunde nachgingen, als erwerbstätig definiert, wurden in der LV-Ost 71 alle Erwerbstätigkeiten erfasst, in der die Zielpersonen sich subjektiv als vorwiegend erwerbstätig einstufen.²²

Um Verzerrungen durch diese unterschiedliche Definitionen zu vermeiden, wurden im Mikrozensus zunächst nur diejenigen als Erwerbstätige gezählt, die ihren überwiegenden Lebensunterhalt durch Erwerbstätigkeit bestritten und keine Lehre absolvierten. Der so definierte Anteil erwerbstätiger Personen ist im Mikrozensus in allen Vergleichsjahren höher als der Anteil der Personen in der LV-Ost 71, die eine Erwerbstätigkeitsepisode im jeweiligen Stichmonat hatten. Wiederum drängt sich der Verdacht auf, dass dieser Befund auf eine höhere Teilnahmbereitschaft Höherqualifizierter an der LV-Ost 71 zurückzuführen ist, da mit einem höheren Bildungsniveau in der Regel längere Ausbildungszeiten verbunden sind, die zu einem verzögerten Eintritt ins Erwerbsleben führen. Wenn die Erwerbsquote nach dem höchsten Schulabschluss differenziert wird, ist der Erwerbstätigenanteil in der LV-Ost 71 bei den Abiturienten in den Jahren 1991 und

22 Alle anderen entlohnten Tätigkeiten wurden als Nebenbeschäftigungen erfasst.

1996 und bei den Personen ohne Abitur in den Jahren 1991 und 1993 signifikant niedriger (vgl. Abbildung 6).²³

Abbildung 6: Erwerbstätigenanteil, differenziert nach Schulbildungsniveau



Chi-Quadrat: ohne Abitur-1991: 7,4***; 1993: 10,9***; 1995: 0,4; 1996: 0,1/ mit Abitur-1991: 7,0***; 1993: 2,3; 1995: 1,6; 1996: 3,9**

Die höhere Erwerbstätigenquote bei den Abiturienten im Mikrozensus 1991 ist darauf zurückzuführen, dass die Absolventen einer Berufsausbildung mit Abitur in der Regel zunächst von ihrem Ausbildungsbetrieb übernommen wurden, später jedoch ein Studium aufnahmen. Dies hängt möglicherweise mit Problemen der retrospektiven Erfassung von Lebensverlaufsdaten zusammen. Rückblickend werden Erwerbsepisoden, wenn sie nur sehr kurz andauerten, seltener berichtet.

Differenziert man die Erwerbstätigenquote nach Geschlecht, stellt sich heraus, dass die Erwerbsbeteiligung bei den Männern in beiden Studien mehr oder weniger kontinuierlich steigt. In der LV-Ost 71 ist diese im Laufe der Zeit zunehmende Erwerbsbeteiligung bei den Frauen ebenfalls festzustellen. Im Mikrozensus sinkt demgegenüber die Erwerbsbeteiligung der Frauen zwischen 1991 und 1993 um 7 Prozent. Dieser Trend ist vermutlich auf die Ausgrenzung von Frauen aus dem ostdeutschen Arbeitsmarkt nach 1991

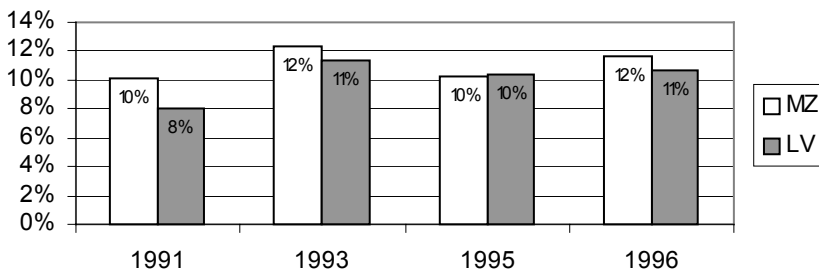
²³ Dass die Erwerbstätigenquote im Mikrozensus 1991 höher liegt als im Mikrozensus 1993 ist insbesondere darauf zurückzuführen, dass die Absolventen einer Berufsausbildung mit Abitur in der Regel zunächst von ihrem Ausbildungsbetrieb übernommen wurden, später jedoch ein Studium aufnahmen.

zurückzuführen. Dass die hohe Erwerbsbeteiligung der Frauen im Jahr 1991 mit den Daten der LV-Ost 71 nicht zu rekonstruieren ist, hängt möglicherweise aber auch mit Problemen der retrospektiven Erfassung von Lebensverlaufsdaten zusammen.

Um die Angaben zur Arbeitslosigkeit zwischen Mikrozensus und LV-Ost 71 vergleichen zu können, wurden im Mikrozensus nur diejenigen erwerbslosen Personen ausgezählt, die angaben, ihren überwiegenden Lebensunterhalt durch Arbeitslosengeld oder -hilfe zu bestreiten. Stellt man diesen Verteilungen die Anteile jener erwerbslosen Personen der LV-Ost 71 gegenüber, die im Stichmonat eine Arbeitslosigkeitsepisode hatten, ist in allen Vergleichsjahren in der LV Ost 71 ein signifikant niedrigerer Arbeitslosenanteil als im Mikrozensus festzustellen. Angesichts der höheren Erwerbsbeteiligung im Mikrozensus sind diese Differenzen unerwartet groß. Sie müssen, da die Zuordnung der arbeitslosen Personen im Mikrozensus eindeutig ist, mit der Konstruktion der Vergleichsvariable in der LV-Ost 71 im Zusammenhang stehen.

Da im Mikrozensus nicht – wie in der LV-Ost 71 – zwischen Bezug von Arbeitslosen- und Unterhaltsgeld unterschieden wird und viele Ostdeutsche im Rahmen von Maßnahmen des Arbeitsamtes Ausbildungen absolvierten, sich im Mikrozensus jedoch offensichtlich als Arbeitslose betrachteten, lässt sich eine bessere Vergleichbarkeit von Mikrozensus und LV-Ost 71 nur durch eine Definitionserweiterung des Arbeitslosenanteils in der LV-Ost 71 gewährleisten. Stellt man den Arbeitslosenanteilen im Mikrozensus den Anteil derjenigen Personen in der LV-Ost 71 gegenüber, die im Stichmonat einen Arbeitslosigkeitsepisode hatten bzw. sich in einer Ausbildung befanden und ihren Lebensunterhalt durch Unterhaltsgeld (vom Arbeitsamt) bestritten, sind die Differenzen zwischen Mikrozensus und LV-Ost 71 nicht mehr signifikant (vgl. Abbildung 7).

Abbildung 7: Anteil arbeitsloser bzw. in Umschulung befindlicher Personen

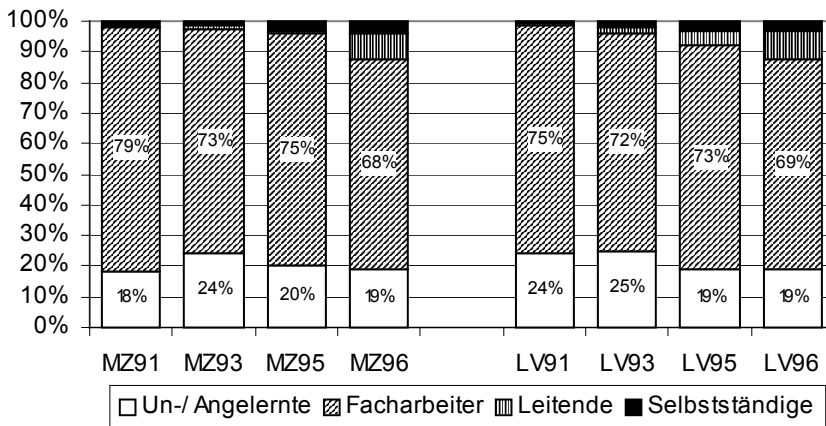


Chi-Quadrat: 1991: 2,4; 1993: 0,4; 1995: 0,0; 1996: 0,5

In der Konsequenz bedeutet das, wenn mit einem sehr differenzierten Bild von Bildungs- und Erwerbsverläufen festgestellt wird, dass die Arbeitslosenquote bzw. die Arbeitslosigkeitsdauer im Vergleich mit anderen statistischen Zahlen weitaus geringer ist, kann dies unter Umständen an den unterschiedlichen Erhebungskonzepten liegen.

Durch den Vergleich der Anteile Erwerbstätiger in verschiedenen beruflichen Stellungen (unqualifizierte, qualifizierte und hochqualifizierte Positionen) lässt sich überprüfen, ob sich der vermutete Bildungseffekt bei den Erwerbstätigen niederschlägt. Dieser Vergleich war allerdings mit dem Problem konfrontiert, dass die Angaben im Mikrozensus 1991, 1993 und 1995 nicht nach den in der LV-Ost 71 verwendeten Kategorisierung vorlagen.²⁴ Ab 1996 wurden fast identische Kategorien verwendet, so dass vor allem hier mögliche Stichprobenverzerrungen durch die von der beruflichen Stellung abhängigen Teilnahmebereitschaft sichtbar werden müssten. Doch gerade im Jahre 1996 sind keine signifikanten Unterschiede zwischen Mikrozensus und LV-Ost 71 hinsichtlich der beruflichen Stellungen festzustellen. Und obwohl bisher ein Bildungseffekt vermutet wurde, erstaunt der im Jahre 1991 erkennbare höhere Anteil Un- und Angelernter in der LV-Ost 71. Konsistent ist dagegen der im Jahre 1995 niedrigere Anteil an Leitenden im Mikrozensus (siehe Abbildung 8).

24 Für diese Jahre wurden aus dem Mikrozensus die Gruppe 4 den "Un-/ Angelernten", die Gruppen 5, 6 und 7 den "Facharbeitern", die Gruppen 8, 9 und 10 den "Leitenden", die Gruppe 2 den "Selbstständigen" zugeordnet. Aus der LV-Ost 71 wurden die differenzierten beruflichen Stellungen 40, 51, 60, und 61 als "Un-/ Angelernte", die 12, 41, 42, 52 und 62 als "Facharbeiter", die 43, 53, 63, 64 und 54 als "Leitende" sowie die 10, 11, 15, 16, 17, 21, 22, 23 und 24 als "Selbstständige" betrachtet.

Abbildung 8: Erwerbstätige: Berufliche Stellung

Chi-Quadrat: 1991: 6,2; 1993: 0,9; 1995: 21,9***; 1996: 0,2

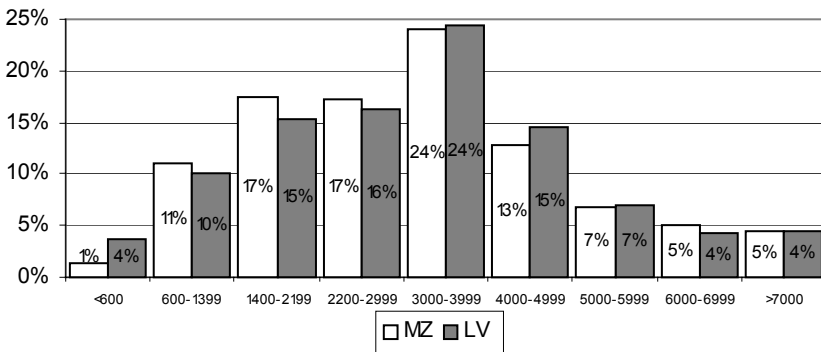
Sowohl bei den Frauen als auch bei den Männern folgen die Abweichungen keinem erkennbaren Muster. Bei den Männern sind sie nur 1995 signifikant und stehen mit dem in diesem Jahr deutlich höheren Anteil von Männern in leitenden Positionen im Zusammenhang. Bei den Frauen beruht die Signifikanz der Abweichungen auf einem deutlich höheren Un- und Angelerntenanteil in der LV-Ost 71, was sich vermutlich auf Zuordnungsschwierigkeiten bei den Befragten zurückführen lässt. Viele Berufsbezeichnungen in der DDR enthielten den Begriff „Facharbeiter“ unabhängig von der mit der Tätigkeit verbundenen beruflichen Stellung, so dass gerade in den ersten Jahren nach der Wiedervereinigung dieses Problem gehäuft bei Befragungen Ostdeutscher auftrat.

2.1.5 Einkommen

Um die finanzielle Situation der Zielpersonen vergleichen zu können, wurde in beiden Studien das monatliche Haushaltsnettoeinkommen herangezogen. In der LV-Ost 71 wurde diese Variable jedoch nur für den Interviewzeitpunkt erfasst, so dass im Folgenden das aktuelle monatliche Haushaltsnettoeinkommen im April 1996 aus dem Mikrozensus gegenübergestellt wird. In der LV-Ost 71 liegt diese Variable in kontinuierlichen Zahlen vor, im Mikrozensus dagegen kategorisiert. Für den Vergleich wurden die Angaben aus der LV-Ost 71 in die Mikrozensus-Klassifizierung übertragen und dann einige Kategorien zusammengefasst. Vergleicht man die Verteilung des aktuellen monatlichen Haushalts-

nettoeinkommens wird deutlich, dass – abgesehen von der niedrigsten Einkommensgruppe – die niedrigeren Einkommensgruppen im Mikrozensus und – abgesehen von den beiden höchsten Einkommensgruppen – die höheren Einkommensgruppen in der LV-Ost 71 stärker vertreten sind (vgl. Abbildung 9).

Abbildung 9: Haushaltsnettoeinkommen 1996 (in DM)



Chi-Quadrat: 1996: 13,474

Dieses Ergebnis war zu erwarten gewesen, da ja in der LV-Ost 71-Stichprobe höher Gebildete stärker vertreten waren als im Mikrozensus. Der Bildungseffekt kommt allerdings (noch) nicht so stark zum Tragen, da sich 1996 ein großer Teil der Hochqualifizierten der 1971er Geburtskohorte in Ausbildung befanden. Deshalb sind wahrscheinlich in der LV-Ost 71 auch deutlich häufiger als im Mikrozensus Haushaltsnettoeinkommen unter 600 DM zu finden. Dass im Mikrozensus die höchsten Einkommensgruppen häufiger vertreten sind als in der LV-Ost 71, kann auf die geringere Auskunftsbereitschaft bzw. Erreichbarkeit von regional äußerst mobilen, viel arbeitenden und sehr gut verdienenden Personen zurückgeführt werden. Diese Personen werden im Mikrozensus aufgrund der bestehenden Auskunftspflicht besser erfasst als in der LV-Ost 71. Insgesamt sind die Unterschiede zwischen dem Mikrozensus und der LV-Ost 71 hinsichtlich des aktuellen monatlichen Haushaltsnettoeinkommens für das Jahr 1996 jedoch als gering zu betrachten.

2.1.6 Zusammenfassung

Um zusammenfassend die Abweichungen zwischen den Randerteilungen des Mikrozensus und der LV-Ost 71 zu beschreiben, werden im Folgenden die univariaten Verteilungen mittels Likelihood-Ratio-Test verglichen (vgl. Tabelle 4).

Tabelle 4: Devianzen der Randverteilungen

diff(-2LogL)	DF	1991	1993	1995	1996
Geschlecht	1	0,72	0,35	0,85	0,44
Bundesland	5	15,00**	5,18	1,47	4,15
Gemeindegröße	2	10,90***	10,55***	0,36	0,33
Familienstand	2	12,52***	10,37***	13,52***	4,88*
Kinder ja/ nein	1	25,70***	9,27***	5,85**	1,97
Anzahl Kinder	2	25,92***	9,54***	8,39**	2,94
Stellung im Haushalt	1	0,06	0,01	0,03	0,50
Höchster Schulabschluss	2	8,06**	3,06	7,17**	7,96**
Erwerbsbeteiligung	1	4,58**	18,28***	4,85**	4,26**
Arbeitslosigkeit	1	2,54	0,43	0,00	1,02
Berufliche Stellung	3	6,38*	0,79	18,03***	0,16
Höchster Ausbildungsabschluss	5				13,86**
Haushaltsnettoeinkommen	19				41,02***

Statistische Signifikanz gekennzeichnet: ***: 1-Prozent-Niveau, **: 5-Prozent-Niveau, *: 10-Prozent-Niveau

Beim Unabhängigkeitsmodell wird davon ausgegangen, dass sich die jährlichen Verteilungen des Mikrozensus und der LV-Ost 71 lediglich hinsichtlich ihrer Stichprobengröße unterscheiden. Die Differenz zwischen den durch dieses Unabhängigkeitsmodell geschätzten und den beobachteten Daten sagt etwas über die Abweichung bezüglich der

Teilnahmewahrscheinlichkeit an der LV-Ost 71 gegenüber dem Mikrozensus aus. Statistisch hoch signifikant sind vor allem Abweichungen, die mit Familie, Bildungsniveau und Erwerbsbeteiligung im Zusammenhang stehen.

2.2 Multivariater Vergleich der LV-Ost 71 mit dem Mikrozensus

Um sicher zu gehen, dass die festgestellten Abweichungen zwischen LV-Ost 71 und Mikrozensus nicht auf ein und dieselbe Verzerrung zurückzuführen sind, werden multivariate Logit-Modelle, die die relative Wahrscheinlichkeit der Teilnahme an der LV-Ost 71 gegenüber der Teilnahme am Mikrozensus schätzen, berechnet. In Tabelle 5 (Seite 36) werden die geschätzten Regressionsgewichte (Beta) aufgeführt. Dabei sind positive Werte ein Hinweis auf eine überproportionale und negative Werte auf eine unterproportionale Beteiligung an der LV-Ost 71 gegenüber dem Mikrozensus.²⁵

Wie oben bereits beschrieben, lassen sich die signifikante Unterrepräsentation der in Städten mit mehr als 500.000 Einwohnern Wohnenden vor allem darauf zurückführen, dass die Wahrnehmung der Einwohnerzahl durch die Befragten nicht mit der amtlich registrierten Einwohnerzahl identisch sein muss. Das Verschwinden der Signifikanz ab 1995 steht mit der Tatsache im Zusammenhang, dass in einigen ostdeutschen Großstädte die Einwohnerzahl trotz Verwaltungsreform unter die 500.000-Marke gefallen ist.

Die Überrepräsentation Kinderloser in der LV-Ost 71 ist einerseits auf eine geringere Teilnahmebereitschaft von Eltern, die sehr früh Kinder bekommen haben, und andererseits auf eine Überrepräsentation Lediger zurückzuführen. Warum Ledige in der LV-Ost 71 überrepräsentiert sind, lässt sich auch durch die Berücksichtigung verschiedener Interaktionseffekte nicht vollständig aufklären. Deutlich wird zwar, dass der Bildungseffekt (Abiturienten sind häufiger ledig) eine gewisse Erklärungskraft für die Überrepräsentation Lediger hat, aber er reicht nicht aus, um das Phänomen wirklich zu beschreiben. Ein bedeutsamerer Grund für diese Abweichung könnte sein, dass – weil bei einer Heirat häufig der Nachname des Partners angenommen wird – Adressen Unverheirateter leichter zu recherchieren waren.

25 Aufgrund von Multikollinearität können natürlich nicht alle Variablen gleichzeitig in das Gesamtmodell aufgenommen werden. Darüber hinaus mussten, weil die Zellenbesetzungen bei Berücksichtigung aller Kategorien sehr klein waren, einzelne Kategorien zusammengefasst werden.

Tabelle 5: Parameter der Teilnahme an LV-Ost 71 vs. am Mikrozensus (Gesamtmodell)

		1991	1993	1995	1996
Konstante		-2,19***	-1,60***	-1,81***	-1,78***
Geschlecht	Mann	-0,01	0,14	0,10	-0,05
Ref. Frau					
Gemeindegröße	> 500.000	-0,58***	-0,65***	-0,23	0,08
Ref. < 20.000	20.000-500.000	0,07	0,02	-0,15	-0,10
Familienstand	ledig	0,73**	0,26	0,50***	0,30*
Ref. sonstiges					
Kinder	Kinder	-0,83***	-0,57***	-0,14	-0,05
Ref. keine Kinder					
Haushalt	eigener	0,42***	0,33***	0,23**	0,38***
Ref. anderer					
Höchster Schulabschluss	max. 8.Klasse	-0,38	0,14	-0,24	-0,23
Ref. 10.Klasse	Abitur	0,13	-0,41***	-0,05	-0,21
Erwerbsstatus	Nichterwerbstätige	0,30**	0,84***	0,48***	0,54***
Ref. Facharbeiter	Arbeitslose	-0,31	-0,49**	-0,15	-0,45**
	Un-/ Angelernte	0,44***	0,08	0,03	0,14
	Leitende	-0,24	-0,01	1,58***	0,09
	Selbstständige	-0,79	-0,13	-0,11	-0,19
Höchster Ausbildungsabschluss	Ohne Abschluss				0,05
Ref. Lehre	In Ausbildung				-0,59**
	Anlernausbildung				-1,08
	Meister/ Fachschule				0,54***
	(Fach)Hochschule				-0,29
Haushaltsnettoeinkommen	< 1000				-0,00
Ref. 3500 bis <5000	1000 bis < 2200				-0,22
	2200 bis < 3500				-0,15
	5000 bis < 6500				0,15
	6500 bis < 8000				0,00
	> 8000				
	diff(-2LogL)	71,0***	82,9***	48,27***	66,47***

Obwohl bei der Deskription keine signifikanten Unterschiede hinsichtlich der Stellung im Haushalt festgestellt werden konnten, wird interessanterweise bei der multivariaten Analyse deutlich, dass Personen in eigenen Haushalten in der LV-Ost überrepräsentiert sind. Allerdings kann für die unterschiedliche Messung der Haushaltsstruktur eine unterschiedliche Interpretation der jeweiligen Wohnsituation verantwortlich sein. So ist

z.B. unklar, ob ein junger Mann, der zwar bei seinen Eltern in der Wohnung lebt, seinen Haushalt aber – weil er dafür einen Obolus zu entrichten hat – als eigenen Haushalt oder den Haushalt der Eltern betrachtet (vgl. Pötter/Rendtel 1993: 269).

Überrepräsentiert sind Nichterwerbstätige, vor allem Studenten²⁶, was sich einerseits auf eine höhere Teilnahmebereitschaft Hochqualifizierter und andererseits auf eine schlechtere Erreichbarkeit Erwerbstätiger zurückführen lässt. Auch die in der LV-Ost festzustellende Unterrepräsentation von Arbeitslosen ist ein Ergebnis, das in der Umfrageforschung bereits länger bekannt ist.

3. Fazit

Um Aufschluss über eventuelle Stichprobenverzerrungen der LV-Ost 71 zu erhalten, wurden in der vorliegenden Untersuchung ausgewählte Randverteilungen der realisierten Stichprobe mit Mikrozensusserhebungen aus den Jahren 1991, 1993, 1995 und 1996 verglichen. Dieser Vergleich konnte eine Reihe von aus der Umfrageforschung bereits bekannten Ergebnissen hinsichtlich der Nichtteilnahme an freiwilligen Befragungen erneut bestätigen. So nahm, wie bereits mehrfach für Lebensverlaufsstudien belegt wurde, auch in der realisierten Stichprobe die Teilnahmebereitschaft mit steigendem Bildungsniveau zu. Nichterwerbstätige beteiligen sich häufiger, Arbeitslose seltener an der LV-Ost 71 als am Mikrozensus. Neu dagegen ist der Befund, dass in der LV-Ost 71 Ledige und Kinderlose überrepräsentiert sind. Vermutet wurde, dass dieser Effekt auf Schwierigkeiten bei der Recherche der aktuellen Adressen zurückzuführen ist, da die Befragten – vor allem die Frauen – nach einer Hochzeit häufig den Nachnamen des Partners annehmen. Wenn sich die Ausgangsstichprobe auf sehr junge Personen richtet und die Zeitspanne zwischen Stichprobenziehung und Befragung relativ groß ist, sind Verheiratete offensichtlich schlechter zu erreichen. Leider konnte, da in der LV-Ost 71 keine Individualdaten über die Nichterreichbarkeit bzw. Teilnahmeverweigerungen zur Verfügung stehen, dieser Frage nicht weiter nachgegangen werden. Es ist nicht auszuschließen, dass sich dieser Effekt in der vorliegenden Untersuchung auch aufgrund der unterschiedlichen Zielpopulation, der abweichenden Fragestellungen oder aufgrund von Operationalisierungsproblemen ergeben haben. Festzuhalten bleibt deshalb, dass das Wissen über die Komplexität von Teilnahmeausfalleffekten noch so gering ist, dass in Bezug auf diese Frage noch weiterer Forschungsbedarf besteht. Allerdings zeigen sich beim Vergleich keine Hinweise auf systematisch variierende Abweichungen, die mit dem retrospektiven, zeitlichen Befragungsabstand korrespondieren.

26 Wenig überraschend ist, dass Nichterwerbstätigkeit mit Abitur hoch korreliert. Der Bildungseffekt äußert sich weniger in einem direkten positiven Effekt des Abiturs, sondern vermittelt über die Nichterwerbstätigkeit.

Wie ist mit den in der LV-Ost 71 gefundenen Effekten umzugehen? Die LV-Ost 71 ist eine retrospektive Längsschnittbefragung, so dass es – aufgrund der eingeschränkten Verfügbarkeit von im Optimalfall monatlichen Hochrechnungsfaktoren für die entsprechende Grundgesamtheit, unmöglich ist, statistische Korrekturverfahren einzusetzen. Vor diesem Hintergrund, ist besondere Vorsicht bei Aussagen geboten, die sich auf die Grundgesamtheit beziehen, denn es können sich bei einfachen deskriptiven Analysen durch den Selektionsbias Verzerrungen der Ergebnisse ergeben. Allerdings ist eine Repräsentativität der Daten nicht unbedingt erforderlich, um verallgemeinerbare Aussagen über die Grundgesamtheit machen zu können; sie stellt jedoch eine Art Idealtypus dar, dem man sich so weit wie möglich nähern sollte. Daher sollten die Nichtteilnehmer so gut wie möglich charakterisiert und bei den empirischen Analysen identifizierbar gemacht werden. Die Abweichungen zwischen LV-Ost 71 und Mikrozensus sind unproblematisch, wenn bei der Untersuchung der Wechselbeziehungen zwischen verschiedenen Variablen der identifizierte Selektionsbias berücksichtigt wird. Die Ergebnisse der vorliegenden Untersuchung legen nahe, in die Analysen der LV-Ost 71 das zum jeweiligen Zeitpunkt erreichte Bildungsniveau, den jeweils aktuellen Familienstand und den Erwerbsstatus zum Untersuchungszeitpunkt einzubeziehen, um die Schätzergebnisse aus den multivariaten Analysen richtig interpretieren zu können.

Literatur

- Blossfeld, H.-P., 1987: Zur Repräsentativität der Sfb-3 Lebensverlaufsstudie. Ein Vergleich mit Daten der amtlichen Statistik. Allgemeines Statistisches Archiv 71: 126-144.
- Born, C./Erzberger, C., 1999: Räumliche Mobilität und Regionalstichprobe: zum Zusammenhang von Regionalität und Repräsentativität in der Lebenslaufforschung (Arbeitspapier des Sonderforschungsbereichs Statuspassagen und Risikolagen im Lebensverlauf – Institutionelle Steuerung und Individuelle Handlungsstrategien, 58). Bremen: Sonderforschungsbereich Statuspassagen und Risikolagen im Lebensverlauf.
- Bundesministerium für Bildung und Wissenschaft (Hrsg.), 1990/91: Grund- und Strukturdaten, 1990/91. Bonn: Bundesministerium für Bildung und Wissenschaft.
- De Leeuw, E. D., 1992: Data quality in mail, telephone, and face to face surveys. Amsterdam: TT-publikaties.
- Erbslöh, B./Koch, A., 1988: Die Nonresponse-Studie zum Allbus 1986: Problemstellung, Design, erste Ergebnisse. ZUMA-Nachrichten 22: 29-44.
- Esser, H., 1986: Über die Teilnahme an Befragungen. In: ZUMA-Nachrichten 18: 38-47.
- Hartmann, P. H./Schimpl-Neimanns, B., 1992: Sind Sozialstrukturanalysen mit Umfragedaten möglich? Analysen zur Repräsentativität einer Sozialforschungsumfrage. KZfSS 44 (Heft 2): 315-340.

INFAS Sozialforschung, 1998: Ostdeutsche Lebensverläufe im Transformationsprozess – Methodenbericht zur Hauptstudie (1971er Kohorte). Bonn: INFAS.

Lichtwardt, B., 2004: Nachrecherchebericht für die Studie "Ostdeutsche Lebensverläufe im Transformationsprozess" (LV-Ost 71). In: Matthes, B./Lichtwardt, B./Mayer, K. U., 2004: Dokumentationshandbuch. Ostdeutsche Lebensverläufe im Transformationsprozess (Materialien aus der Bildungsforschung 76), Teil 2: Dokumentation verwendeter Materialien. Berlin: Max-Planck-Institut für Bildungsforschung: 93-141.

Matthes, B./Lichtwardt, B., 2004: Editionsbericht für die Lebensverlaufsstudie Ost, Geburtskohorte 1971. In: Matthes, B./Lichtwardt, B./Mayer, K. U., 2004: Dokumentationshandbuch. Ostdeutsche Lebensverläufe im Transformationsprozess (Materialien aus der Bildungsforschung 76), Teil 2: Dokumentation verwendeter Materialien. Berlin: Max-Planck-Institut für Bildungsforschung: 27-92.

Reimer, M., 2001: Die Zuverlässigkeit des autobiographischen Gedächtnisses und die Validität retrospektiv erhobener Lebensverlaufdaten. Kognitive und erhebungspragmatische Aspekte (Materialien aus der Bildungsforschung, 71). Berlin: Max-Planck-Institut für Bildungsforschung.

Rendtel, U./Pötter, U., 1993: Über Sinn und Unsinn von Repräsentativitätsstudien. Allgemeines Statistisches Archiv 77: 260-280.

Schimpl-Neimanns, B./Müller, J., 2001: Zur Berechnung des Stichprobenfehlers im Mikrozensus. Mannheim: ZUMA.

Schnell, R., 1997: Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen. Opladen: Leske+Budrich.

Staudenmaier, B./Granato, N./Schmidt, S., 2000: Vergleich der Mikrozensen 1995 und 1996. ZUMA-Technischer Bericht 2000/08. Mannheim: ZUMA.

Korrespondenzadresse

*Britta Matthes
Max-Planck-Institut für Bildungsforschung
Lentzealle 94
D-14195 Berlin
Telefon: 030/ 82406295
email: matthes@mpib-berlin.mpg.de*

SELEKTIVES ANTWORTVERHALTEN BEI FRAGEN ZUM DELINQUENTEN HANDELN*

Eine empirische Studie über die Wirksamkeit der ‚sealed envelope technique‘ bei selbst berichteter Delinquenz mit Daten des ALLBUS 2000

ROLF BECKER & RALPH GÜNTHER

Im ALLBUS 2000 wurde die ‚sealed envelope technique‘ (SET) eingesetzt, um zuverlässige wie gültige Antworten zum delinquenten Handeln zu erhalten. Im vorliegenden Beitrag gehen wir der Frage nach, warum trotz zugesicherter Anonymität einige der Befragten das Selbstaussfüllen des vertraulichen Fragebogens zum eigenen delinquenten Verhalten verweigern. In theoretischer Hinsicht wird aus Sicht eines werterwartungstheoretischen Erklärungsansatzes angenommen, dass Personen bei Interviews an sozialer Anerkennung und Vermeiden von Missbilligung durch den Interviewer interessiert sind und diejenige Antwortreaktion auswählen, um diese Ziele zu optimieren. Wenn für Befragte die Fragen zum eigenen delinquenten Handeln irrelevant sind oder die SET unverständlich geblieben ist, dann verweigern sie eher die Antworten als Personen, die sowohl die Fragen und die SET verstanden als auch entsprechende Antworten auf Fragen zur eigenen Delinquenz haben. Zum anderen wird davon ausgegangen, dass Interview- und Interviewereffekte auch bei zugesicherter Anonymität zur Antwortverweigerung führen können. Mit Daten des ALLBUS können diese Annahmen weitgehend bestätigt werden. Schließlich wird gezeigt, dass dieses selektive Antwortverhalten zu verzerrten Schätzergebnissen bei Modellen zu Determinanten des beabsichtigten delinquenten Handelns wie Steuerhinterziehung oder „Schwarzfahren“ führen kann. Bei heiklen und unangenehmen Fragen zum delinquenten Handeln wäre die postalische Befragung ein effizienter Ausweg, um dieses methodische Problem zu lösen. Für die endgültige Klärung dieser Sachverhalte

* Für hilfreiche Kommentare danken wir Guido Mehlkop, Franziska Rudolph und besonders den anonymen Gutachtern der Zeitschrift.

benötigen wir mehr Informationen über die Mechanismen und Prozesse, die zur Verweigerung von Antworten führen.

In the German General Survey 2000 (ALLBUS) the so-called sealed envelope technique was used to collect information about the individuals' self-reported delinquency. The article looks in particular at why respondents refuse to fill out this confidential questionnaire in spite of the guaranteed anonymity. From a theoretical perspective of subjective expected utility, it could be assumed that respondents are interested in maximizing benefits and avoiding social costs in the interview situation. Therefore, they respond to questions in a manner to realise their interests in the optimal way. At the same time, individual respondent characteristics related to their understanding of the questions or the sealed envelope technique, as well as interviewer characteristics and aspects of the interview situation could lead to refusals on sensitive questions. The ALLBUS 2000 data confirm these hypotheses. The selectivity of self-reported delinquency on things such as fare-dodging and tax evasion also resulted in biased model estimators of determinants of anticipated future delinquency. Mail surveys could be one way to improve data quality in self-reported acts of delinquency. Before firm conclusions can be drawn, however, more empirical data is needed on the processes and mechanisms behind respondents refusing to answer questions about delinquency.

1. Einleitung

Eine methodische Schwierigkeit von Bevölkerungsumfragen wie dem ALLBUS ist die *Erhebung gültiger wie zuverlässiger Informationen über delinquentes Handeln der Befragten*. Um Antwortverweigerungen und verzerrte Antworten zur eigenen Delinquenz zu vermeiden, sind besondere Vorkehrungen bei der Befragung zu treffen. Als besonders sinnvoll hat sich bei ‚face-to-face‘-Interviews ein vertrauensbildendes Verfahren wie die ‚sealed envelope technique‘ (SET) erwiesen (De Leeuw 2001; Sudman und Bradburn 1974, 1982). Es wurde auch für den Themenbereich ‚Sanktion und abweichendes Verhalten‘ angewendet, der ein Schwerpunktthema des ALLBUS im Jahre 2000 war. Um den Befragten die Anonymität zu garantieren, wurden sie gebeten, Fragen zur früheren und zukünftigen Delinquenz – zwar im Beisein des Interviewers, aber ohne dessen Einsicht in die Antworten der Befragten – selbst auf einem gesonderten Fragebogen schriftlich zu beantworten und diesen anschließend in einem verschlossenen Umschlag dem Interviewer

zu überreichen.¹ Trotz dieser Vorgehensweise gibt es beim ALLBUS 2000 dennoch Personen, die den vertraulichen Fragebogen *nicht* ausgefüllt haben. Die Verweigerungsquote beträgt 12,5 Prozent, und wenn über eine Disproportionalitätsgewichtung die disproportionale Stichprobenziehung für Ostdeutschland berücksichtigt wird, liegt die Verweigerungsquote bei 13,2 Prozent.

Im vorliegenden Beitrag gehen wir der Frage nach, *warum* einige der Befragten trotz zugesicherter Anonymität das Ausfüllen des vertraulichen Fragebogens zum eigenen delinquenten Verhalten verweigern. Sollte sich herausstellen, dass es sich dabei um systematische Ausfälle handelt, dann wird zu überprüfen sein, *ob* die selektiven Antwortverweigerungen zu verzerrten Ergebnissen führen, wenn Prozesse des delinquenten Handelns selbst untersucht werden (vgl. Hindelang et al. 1979). Bei multivariaten Schätzungen müssten dann möglicherweise ‚selectivity bias‘ ausgeglichen werden.

1 Im ALLBUS 2000 erfolgte die Datenerhebung mittels computerunterstützter persönlicher Interviews (CAPI). Für Fragen zum delinquenten Handeln der Befragten hingegen kam die SET zum Einsatz: Die Interviewer sollten, so die explizite Intervieweranweisung, die Befragten darum bitten, einen separaten Fragebogen mit den Fragen zur eigenen Delinquenz selbständig schriftlich auszufüllen („Selbstaussfüller“). Damit der Interviewer die Antworten nicht einsehen kann, sollten die Befragten den Fragebogen in einen Umschlag stecken, diesen verschließen und erst dann dem Interviewer übergeben. Im Falle, dass die Befragten zögern, die Fragen zu beantworten, oder Bedenken äußern, sollten die Interviewer darauf hinweisen, dass die Antworten vertraulich behandelt werden und die statistischen Auswertungen nach den Bestimmungen des Datenschutzgesetzes erfolgen. Insgesamt waren zwei Anonymitätszusicherungen vorgesehen. Um Antwortverweigerungen in größerem Ausmaß zu vermeiden, standen lediglich Gesetzesverstöße zur Diskussion, die in der Regel mit geringen Geldstrafen und nur in schwerwiegenden Fällen mit höheren Geldstrafen oder Gefängnisstrafen sanktioniert werden. So wurden die Respondenten danach befragt, ob und wie oft sie in der Vergangenheit bereits Delikte wie etwa Steuerhinterziehung, Kaufhausdiebstahl, Schwarzfahren oder Trunkenheit am Steuer begangen haben, und ob sie diese Gesetzesübertritte in Zukunft begehen würden. Beim Pretest des ALLBUS 1990, in dem für das gleiche Thema bereits die SET verwendet wurde, berichteten die Interviewer über keine nennenswerten Probleme bei der Anwendung dieses Verfahrens oder bei der Durchführung eines persönlichen Interviews über delinquentes Verhalten. So gab es tendenziell positive Antworten zum eigenen delinquenten Verhalten, wenn die SET verwendet wurde; aber für die einzelnen Delikte wie Schwarzfahren, Steuerbetrug, Ladendiebstahl oder Trunkenheit am Steuer gab es keine signifikanten Unterschiede zwischen den beiden Erhebungsmethoden. Am größten war der Unterschied bei der Steuerhinterziehung. Angesichts dieser Ergebnisse entschied man sich bei der Haupterhebung für die SET (Wasmer et al. 1991: 11-12).

2. Theorie und Hypothesen

Erklärung des Antwortverhaltens bei Fragen zum delinquenten Handeln

Um die aufgeworfene Frage beantworten zu können, wird ein kohärentes Aussagesystem benötigt, aus dem empirisch überprüfbare Hypothesen über das Antwortverhalten bei heiklen, unangenehmen oder bedrohlichen Fragen wie etwa zum delinquenten Handeln abgeleitet werden können. Solch einen, aber bislang für diese Problematik noch nicht empirisch angewendeten Ansatz schlägt Esser (1974, 1986) vor. Aufrichtige und wegen sozialer Erwünschtheit verzerrte Antworten lassen sich nach Esser (1986) „als kombiniertes Resultat von Motiven, Bedürfnissen und Bewertungen einerseits (sei als stabiles Bedürfnis nach sozialer Anerkennung, als Konformität zu Rollenvorstellungen oder als Versuch, irgendein anderes, mit dem Antwortverhalten subjektiv verbundenes Ziel zu erreichen) und gewissen Erwartungen über den Zusammenhang einer Antwort mit gewissen Konsequenzen erklären“ (ders.: 319). Diese theoretischen Prämissen können, so Esser (1986), in eine *allgemeinen Theorie des situationsorientierten Handelns* integriert werden, wonach Befragte die ihnen vorstellbaren Handlungsalternativen wählen, die am ehesten in der von ihnen subjektiv wahrgenommenen Interviewsituation (z.B. Frageinhalt und Antwortvorgaben sowie Anwesenheit Dritter und perzipierte Interviewereigenschaften) geeignet scheinen, bestimmte Ziele (etwa konfliktfreie Abwicklung des einmal begonnenen Interviews) zu realisieren (*Wert-Erwartungstheorie*). In theoretischer Hinsicht wird das Antwortverhalten von Befragten als das Ergebnis einer subjektiven Kosten-Nutzen-Entscheidung zwischen unterschiedlichen Handlungsalternativen wie etwa aufrichtige Antwort mit dem „wahren Wert“, sozial erwünschte Antwort, zufällige Antwort oder Antwortverweigerung modelliert (Esser 1986: 320): „Diese Entscheidung erfolgt einerseits auf der Grundlage einer Orientierung an den Präferenzen, Zielsetzungen und normativen Einbindungen der Personen und zweitens vor dem Hintergrund der Perception bzw. Kognition der Situation und den damit jeweils vorliegenden Ambiguitäten, Risiken und Möglichkeiten“. Heikle, unangenehme oder bedrohliche Fragen stellen demnach einen *Spezialfall des Problems sozialer Erwünschtheit* dar. Das Antwortverhalten der Befragten hängt dann von der *angenommenen Überprüfbarkeit der Angaben* und von der *vermuteten Vertraulichkeit der Situation und Datenbehandlung* ab.

Der *subjektiv erwartete Nutzen* von aufrichtigen oder sozial erwünschten Antworten kann einerseits in der sozialen Anerkennung als kooperative Befragungsperson, Erfüllung einer Höflichkeitsnorm gegenüber Fremden, Loyalität zu einer öffentlichen Einrichtung oder Unterstützung der Wissenschaft begründet sein. Andererseits kann die Antwortverweigerung dann Nutzen erbringen, wenn beispielsweise Konformität mit Normen von bestimmten (delinquenten) Bezugsgruppen oder grundsätzliche Abneigung gegenüber wissen-

schaftlichen Befragungen erwartet wird. *Kosten* sind beispielsweise mit der subjektiv erwarteten Missbilligung von bestimmten Antworten durch den Interviewer verbunden. Des Weiteren führt Esser (1986) *Transaktionskosten* an, die für die befragte Person durch ihre Unsicherheit im Umgang mit einer fremdartigen Situation wie dem Interview und durch Verständigungsprobleme während des Interviews entstehen können. Weitere Kosten sind gerade bei bedrohlich wirkenden Fragen mit den Befürchtungen von einer Deanonymisierung des Befragten verbunden. In diesem Falle sind vertrauensbildende Maßnahmen wie etwa die Anwendung der ‚sealed envelope technique‘ notwendig, „damit der Befragte wegen dem Vertrauen in die Anonymitätssicherung des Interviewers keinen Grund zu derartigen Befürchtungen hat“ (Esser 1986: 321). Da bei der SET zumindest der Interviewer die Antworten der befragten Person nicht erfährt, – die befragte Person füllt ohne verbale Äußerungen von Antworten einen separaten Fragebogen eigenständig aus, den sie verschlossen in einem Umschlag dem Interviewer übergibt – würde man dann eine Interviewsituation herstellen, in der Antwortreaktionen weder missbilligt noch unterstützt werden. Systematische Antwortverzerrungen müssten dann unwahrscheinlich werden. Wird jedoch vom Befragten die Anonymitätssicherung generell von vornherein als unzureichend bewertet, etwa wegen der Anwesenheit des Interviewers oder von dritten Personen, dann werden mit Sicherheit die Antworten zum eigenen delinquenten Handeln ausbleiben.

Jedoch sind wahrgenommene Bedrohlichkeit der Fragen und unzureichende Zusicherung von Anonymität sowie damit verbundenen Kosten- und Nutzen-Erwartungen nicht die einzigen Gründe dafür, dass bei Fragen zur eigenen Delinquenz (aufrichtige) Antworten verweigert werden. Ausgehend von den Kombinationen, die sich aus der Handlungstendenz für eine „wahre“ Antwort einerseits und der Handlungstendenz zu einer sozial erwünschten Antwort andererseits ergeben, lassen sich von den vier möglichen Typen der Befragungssituation folgende drei Konstellationen des Interviews ableiten, bei denen Antwortverweigerungen trotz zugesicherter Anonymität auftreten können.² Diese Typen von Befragungssituationen bezeichnet Esser (1986: 327-330) als *Indifferenz*, *Situationseffekt* und *Inkonsistenz*. Bei Indifferenz und Inkonsistenz lässt sich das Befragtenverhalten

2 Die Handlungstendenz für eine „wahre“ Antwort ergibt sich nach Esser (1986: 327) aus dem Produkt von der Intensität, dass eine befragte Person eine Meinung oder Einstellung zu einer Frage („wahrer Wert“) hat, und ihrer subjektiven Erwartung, dass eine bestimmte Antwort mit dem „wahren Wert“ oder der personalen Identität der befragten Person korrespondiert. Voraussetzung dafür, dass eine Person diese subjektive Erwartung entwickeln kann, ist eine eindeutig formulierte Frage, die von der befragten Person auch verstanden wird. Die Handlungstendenz zu einer sozial erwünschten Antwort wird durch die Intensität situationaler Bedürfnisse oder Interessen (etwa subjektiv wahrgenommenes Bedrohungspotential der Frage) und der subjektiven Erwartung, dass eine bestimmte Antwort zu sozial erwünschten Konsequenzen führt, bestimmt.

in Bezug auf heikle oder bedrohliche Fragen zur Delinquenz, sofern keine zusätzlichen Randbedingungen benannt werden, nicht eindeutig vorhersagen, während bei der Dominanz von Situationseffekten sozial erwünschte Antworten sehr wahrscheinlich sind:

(1) *Indifferenz*: Wenn die befragte Person keine Einstellung oder Meinung zum befragten Thema entwickeln kann und ihr zudem unklar bleibt, was sie überhaupt auf die Frage antworten soll, dann wird in dem Fall, dass *keine* situativen Anreize für bestimmte Antworttendenzen vorliegen und vom Befragten *keine* Konsequenzen für eine der möglichen Antwortreaktionen erwartet werden, wegen *Indifferenz* eine zufällige Antwortreaktion wahrscheinlich. Eine der möglichen Antwortreaktionen wäre ‚item nonresponse‘. Vor allem dürfte ‚item nonresponse‘ bei nicht vorhandenen Interview- und Interviewereffekten auftreten, wenn – abgesehen von dem recht unwahrscheinlichen Fall, dass sich die befragte Person niemals mit Delinquenz beschäftigt hat, ihr das Thema irrelevant erscheint und mit dem Thema überhaupt nichts anfangen kann – die Fragen schwer verständlich oder gar unverständlich formuliert sind.

Zwar spricht dagegen, dass die entsprechenden Fragen im ALLBUS 2000 nachvollziehbar formuliert sind und sich im Pretest bewährt haben.³ Allerdings ist nicht auszuschließen, dass es sich für *Ausländer mit für das Interview unzureichenden Deutschkenntnissen* ebenso wie für *Personen mit niedrigem Bildungsniveau* als schwierig erweisen kann, die Interviewsituation insgesamt oder die Fragen zur Delinquenz einschließlich der Erläuterungen zur ‚sealed envelope technique‘ eindeutig zu verstehen (vgl. Esser 1974: 124).

Des Weiteren könnte es für ältere Personen problematisch sein, geringfügige, aber in der Vergangenheit weit zurückliegende Gesetzesübertretungen adäquat zu erinnern (vgl. Becker 2001). Liegt bei älteren Befragten eine ausgeprägte Tendenz für aufrichtige Antworten vor, dann wären bei Fragen zur Häufigkeit früherer Straftaten neben verzerrten Antworten auch *Antwortverweigerungen wegen Erinnerungsproblemen* wahrscheinlich. Dies dürfte vor allem dann der Fall sein, wenn die Antwortkategorie „Weiß nicht“ nicht

3 So lautet beispielsweise der Originalwortlaut der Frage zu zukünftigen Gesetzesübertretungen wie etwa das „Schwarzfahren“ folgendermaßen: „Unabhängig davon, ob Sie die genannten kleineren Gesetzesübertretungen in der Vergangenheit tatsächlich schon einmal begangen haben oder nicht: Können Sie sich vorstellen, dass Sie in Zukunft so etwas unter Umständen (wieder) tun würden, oder würden Sie so etwas unter keinen Umständen (wieder) tun? Bitte kreuzen Sie bei jeder der vier genannten kleineren Gesetzesübertretungen Ihre Antwort auf diese Frage an. Öffentliche Verkehrsmittel benutzen, ohne einen gültigen Fahrausweis zu besitzen“. Folgende Antwortkategorien waren vorgegeben: „1. Ja, würde ich unter Umständen (wieder) tun. 2. Nein, würde ich unter keinen Umständen (wieder) tun.“

vorgegeben ist (vgl. De Leeuw 2001).⁴ Dies ist beim ALLBUS 2000 der Fall, obwohl diese Antwort eine zulässige auf die Retrospektivfragen wäre. Zudem werden adäquate Erinnerungsleistungen – betrachtet man die in *Fußnote 4* aufgezeigten Antwortkategorien für die Retrospektivfrage – wegen fehlender kognitiver Anker („cues“) erschwert (Sudman et al. 1996; vgl. Schwarz et al. 1994: 191-192).

Schließlich können *subkulturelle Normen* das Antwortverhalten in Bezug auf das delinquente Verhalten der befragten Person beeinflussen. So kann angenommen werden, dass Personen entsprechend ihrer „kulturellen Identität“ auch von nicht begangenen Delikten berichten, um mit den subkulturellen Normen ihrer delinquenten Bezugsgruppe und somit ihrer kulturellen Identität überein zu stimmen. Hingegen ist nicht auszuschließen, dass vorbestrafte Bürger dazu neigen, Straftaten zu verschweigen oder wegen kognitiver Dissonanz oder Befürchtungen jegliche Antworten auf selbst begangene Delikte verweigern (Diekmann 1980: 47-49). Im Unterschied zum vorherigen Fall sind dann kategorische Antwortverweigerungen wahrscheinlicher.

Insgesamt ist bei dieser Interviewsituation zu erwarten, dass Ausländer, Befragte mit niedrigerem Bildungsniveau, ältere Befragte, Vorbestrafte und Befragte, welche subkulturelle Normen delinquenter Bezugsgruppen internalisiert haben, mit einer vergleichsweise höheren Wahrscheinlichkeit zu ‚item nonresponse‘ neigen (Sudman und Bradburn 1974).

(2) *Situationseffekt*: Ist die Handlungstendenz für valide Antworten gering und dominieren daher Situationseffekte, dann tritt auch bei zugesicherter Anonymität ‚item nonresponse‘ auf. Beispielsweise könnte die *Anwesenheit einer dritten Person* den Befragten als bedrohlich erscheinen, so dass bei Fragen zur eigenen Delinquenz Antwortverweigerungen recht wahrscheinlich sind. Denn gerade bei Anwesenheit dritter Personen sind nach Esser (1974) die Bedingungen für Antwortverweigerungen in idealtypischer Weise erfüllt: Je höher das Bedrohungspotential von Fragen erscheint und je größer die subjektiv erwartete Wahrscheinlichkeit ist, dass eine bestimmte Antwort zu sozial unerwünschten oder gar unerwarteten Konsequenzen führt oder führen wird, desto wahrscheinlicher treten wegen der als Kosten befürchteten Reaktionen auch Antwortverweigerungen auf.

4 Der Wortlaut für die Retrospektivfrage zur Steuerhinterziehung lautet: „Wie Sie wissen, begehen viele Bürger hin und wieder eine kleinere Gesetzesübertretung. Im Folgenden sind vier solcher kleineren Gesetzesübertretungen genannt. Bitte kreuzen Sie bei jeder dieser vier Verhaltensweisen an, wie oft Sie in Ihrem Leben so etwas schon getan haben. Falsche Angaben bei der Einkommensteuererklärung oder beim Lohnsteuerjahresausgleich gemacht, um weniger Steuern zahlen zu müssen.“ Folgende Antwortkategorien wurden vorgegeben: „1) Noch nie – 2) 1mal – 3) 2 bis 5mal – 4) 6 bis 10mal – 5) 11 bis 20mal – 6) mehr als 20mal“.

Ferner wirken sich (sichtbare) *Interviewereigenschaften* auf die Bereitschaft aus, auf Fragen zum delinquenten Handeln zu reagieren. So gilt die *Stereotypisierung* als eine der Ursachen für Antwortverweigerungen. Alltagsvorstellungen – wie beispielsweise, dass Delinquenz „Männersache“ sei – sind Gründe dafür, dass Befragte zu sozial erwünschten Antworten neigen, wenn der Interviewer männlich ist. Nach Esser (1986: 326) können sich bei gegebenen situationalen Interessen der Befragten Handlungstendenzen zu sozial erwünschten Antworten wegen sichtbaren Interviewereigenschaften *und* Stereotypisierungen nur dann manifestieren, wenn die Befragten annehmen (müssen), dass ihre Antworten auf heikle oder bedrohliche Fragen dem Interviewer, den anwesenden dritten Personen oder der Öffentlichkeit bekannt werden. Wird jedoch per SET die Anonymität der Antworten und Vertraulichkeit der Daten zugesichert, dann ist demnach davon auszugehen, dass keine sozial erwünschten Antwortreaktionen auftreten. In diesem Falle wäre zu erwarten, dass sich das Geschlecht des Interviews *nicht* auf das Antwortverhalten der Befragten auswirkt.

Systematische Antwortverzerrungen hängen oftmals mit dem *Alter des Interviewers* zusammen. Mit höherem Alter des Interviewers können Assoziationen bei der befragten Person verknüpft sein, dass ältere Interviewer delinquentes Handeln eher missbilligen als jüngere Interviewer. In diesem Falle würden bei älteren Interviewern Antwortverweigerungen eher auftreten als bei jüngeren Interviewern. Dabei ist zu bedenken, dass möglicherweise das Alter mit dem äußeren Erscheinungsbild des Interviewers korrespondiert, das in Bezug auf Delinquenz sozial erwünschte Antworttendenzen fördert.

Sicherlich trägt auch das *Handeln des Interviewers* selbst – beabsichtigt oder unbeabsichtigt – dazu bei, dass Befragte den vertraulichen Fragebogen nicht ausfüllen. Die Vermittlung seiner Einstellungen zu delinquentem Handeln sowie seiner Erwartungen an die befragte Person wären ebenso Beispiele dafür, dass die Art und Weise der Erläuterung der SET in methodisch unerwünschtem Antwortverhalten resultiert. Mit dem Mangel an Erfahrung als Interviewer dürften solche verzerrende Effekte wahrscheinlicher sein, die schließlich zur Antwortverweigerung führen.

Insgesamt ist für diese Interviewkonstellation zu erwarten, dass *Situationseffekte* – die *Anwesenheit dritter Personen beim Interview* und das *höhere Alter des Interviewers* und die *mit seiner Erfahrung als Interviewer korrelierenden Interviewereffekte* – zur Verweigerung von Antworten zum delinquenten Verhalten beitragen. Die mit dem *männlichen Geschlecht des Interviewers* konfundierte Stereotypisierung dürfte sich bei Verwendung der ‚sealed envelope technique‘ kaum auf das Antwortverhalten der Befragten auswirken.

(3) *Inkonsistenz*: Im besonderen Fall, in welchem sowohl eine Tendenz zur validen Antwort als auch eine Tendenz zu sozial erwünschten Antworten vorhanden ist, wird von

Esser (1986: 329) nicht ausgeschlossen, dass bereits eine *geringfügige Veränderung in der externen Interviewsituation* (z.B. Eindeutigkeit des Fragestimulus, sichtbare Interviewermerkmale oder Anwesenheit Dritter) zu inkonsistentem Antwortverhalten führen kann. Valide Antworten zu delinquentem Handeln sind in unserem Fall dann zu erwarten, wenn – abgesehen davon, dass mit dem Befragungsinstrument alles stimmt und die befragte Person eine Einstellung zum Thema „Sanktion und abweichendes Verhalten“ hat – die Person glaubt, eine passende Antwort auf die Fragen zu haben. Wenn die befragte Person eher annimmt, dass eine bestimmte Antwort zu sozial erwünschten Konsequenzen führt, dann sind eher sozial erwünschte oder verzerrte Antworten zu erwarten.

In Situationen, in denen das Antwortverhalten vornehmlich von diesen subjektiven Erwartungen abhängt, kann es bei heiklen Fragen sinnvoll sein, externe Einflüsse beim Interview auszuschalten, um valide und zuverlässige Antworten zu fördern. Die Zusicherung der Anonymität wäre allerdings nur dann sinnvoll, wenn es überhaupt eine „wahre“ Antworttendenz beim Befragten gibt (Esser 1986: 329). Der Einsatz der SET würde daher – wie im ersten Fall der durch indifferentes Befragtenverhalten geprägten Befragungssituation – nicht immer zum gewünschten Erfolg führen. Insgesamt sind in durch Inkonsistenz geprägten Befragungssituationen Antwortverweigerungen kaum vorhersagbar.

Auswirkungen des selektiven Antwortverhaltens auf statistische Analysen

Ob die zuvor diskutierten systematischen Antwortverweigerungen zu verzerrten Ergebnissen multivariater Schätzungen führen, soll anhand der *ökonomischen Theorie der Kriminalität* von Gary S. Becker (1968) überprüft werden. Nach Becker (1968) stellen Gewinne B aus einer Straftat und zu erwartenden Strafen C sowie die Wahrscheinlichkeit p , bei der Straftat entdeckt und dafür belangt zu werden, die grundlegenden Bedingungen dafür dar, ob eine Person eine Straftat S begeht oder nicht: $P[S] = B - p \times C$. Demnach sind Straftaten desto wahrscheinlicher, je größer die erwarteten Gewinne, je geringer die negativen Sanktionen und je geringer die Entdeckungswahrscheinlichkeiten sind – wenn also gilt: $B > p \cdot C$.

In Anlehnung an das von Heckman (1979) vorgeschlagene Verfahren zur Korrektur von ‚sample selection bias‘, wird das Grundmodell von Becker (1968) mit einer instrumentellen erklärenden Variablen, der *Inklusionswahrscheinlichkeit* λ , ergänzt: $P[S] = B - p \cdot C + \lambda$. Diese Inklusionswahrscheinlichkeit ist die bedingte Wahrscheinlichkeit, dass eine Person den vertraulichen Fragebogen in Abhängigkeit der zuvor diskutierten Individualmerkmale und Interview- und Interviewereffekte ausgefüllt hat. Somit dient diese Instrumentalvariable bei den weiterführenden Analysen des delinquenten Handelns als *Gradmesser für Selektivitätsverzerrung*. Treten ‚selectivity bias‘ auf, dann sind die statistischen Befunde zum selbst berichteten delinquenten Handeln als problematisch einzustufen.

3. Daten, abhängige Variable und statistisches Verfahren

Datenbasis

Für die Analysen werden Querschnittsdaten der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (*ALLBUS*) aus dem Jahre 2000 herangezogen. Es wurden 3.138 Personen deutscher Staatsangehörigkeit befragt, die in Privathaushalten leben und bis zum Befragungszeitpunkt das 18. Lebensjahr vollendet haben. Wegen eines Fragebogen-Splittings wurden nur 1.605 Befragte zur eigenen Delinquenz befragt. Für die Überprüfung der Hypothesen können daher nur Auswertungen mit diesen Befragten des ersten Splits gemacht werden. Nach der Disproportionalitätsgewichtung umfasst die Analysestichprobe 1.613 Personen.

Abhängige Variable und statistisches Verfahren

Die *abhängige Variable* ist die Weigerung der Befragten, den vertraulichen Fragebogen zu ihrem delinquenten Handeln auszufüllen. Um die Verteilung der Wahrscheinlichkeiten für ‚item nonresponse‘ zu schätzen, wird für multivariate Analysen des Antwortverhaltens die *binäre logistische Regression* herangezogen (Urban 1993). Die Beziehung zwischen den bedingten Wahrscheinlichkeiten für das Ausfüllen des vertraulichen Fragebogens und der erklärenden Variablen wird durch folgende, die logistische Wahrscheinlichkeitsverteilung wiedergebende Gleichung beschrieben:

$$\ln \left(\frac{p(\text{Antwort})}{p(\text{Antwortverweigerung})} \right) = \beta_0 + \sum \beta_i \cdot x_i \quad \text{bzw.}$$

$$\frac{p(\text{Antwort})}{p(\text{Antwortverweigerung})} = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}}$$

wobei β_0 die Regressionskonstante darstellt und β_i die zu schätzenden Regressionsgewichte für die Prädiktoren x_i sind.

Methodische Schwierigkeiten beim empirischen Vorgehen

Ein entscheidender Vorteil des theoretischen Ansatzes zur Erklärung des Befragtenverhaltens von Esser (1974, 1986) ist die Integration der Gründe und Bedingungen für systematisch verzerrte Antworten und Antwortverweigerungen in ein kohärentes Aussagesystem. Jedoch weist Diekmann (1995: 380) auf eine methodische Schwierigkeit dieses Ansatzes im konkreten Anwendungsfall hin. So sind auf der einen Seite die Bewertungen von

Kosten und Nutzen schwierig zu identifizieren und auf der anderen Seite dürfte es ebenso schwierig sein, die subjektiven Erwartungen von Befragten zu operationalisieren.

Da im ALLBUS 2000 keine Informationen vorliegen, die es erlauben, die Kosten- und Nutzenkomponenten sowie die subjektiven Erwartungen der Befragten zu identifizieren und zu operationalisieren, verbleibt noch die Möglichkeit, diese indirekt über „Brückenannahmen“ herzuleiten oder über empirische Korrelate zu messen, die sich aus der Theorie des Befragtenverhaltens ableiten lassen. Solch eine *Indizienanalyse* bleibt unvollständig, weil die entscheidenden Mechanismen der Kosten-Nutzen-Erwägung *nicht de facto* gemessen werden, sondern ihre Existenz und Wirkungsweise aus theoretischen Zusatzannahmen und empirischen Messungen von Korrelationen zwischen abhängiger und erklärender Variablen geschlossen und für plausibel gehalten werden (vgl. Hedström und Swedberg 1998). Diese Einschränkung ist bei der Interpretation der empirischen Befunde implizit zu berücksichtigen.

4. Empirische Ergebnisse

Determinanten des Antwortverhaltens bei Fragen zum delinquenten Handeln

In *Tabelle 1* sind die Ergebnisse der Modellschätzungen für das Ausfüllen des anonymen „Selbstaussfüllers“ als abhängiger Variablen dokumentiert. Wie theoretisch erwartet, beantworten *Personen mit deutscher Staatsangehörigkeit* eher die Fragen zum eigenen delinquenten Verhalten als Ausländer (*Modell 1* oder *3*). Neben den befürchteten Konsequenzen, die mit dem Ausländerstatus zusammenhängen, sind – da das *Bildungsniveau der Befragten* kontrolliert wird – offensichtlich Sprachschwierigkeiten verantwortlich dafür.⁵

Entgegen der theoretischen Annahme, dass minder gebildete Personen eher Probleme haben als höher gebildete, die Interviewsituation oder das Prozedere der SET zu verstehen, liegen *keine Bildungseffekte* für das Ausfüllen des anonymen Fragebogens vor. Zwar beantworten Befragte mit mindestens Fachhochschulreife diese Fragen eher als Befragte ohne Schulabschluss oder mit einem Haupt- oder Realschulabschluss. Jedoch ist dieser Befund allenfalls auf dem 10-Prozent-Niveau gesichert.

5 Des Weiteren wurde bei explorativen Datenanalysen auch die Wechselwirkung zwischen Staatsangehörigkeit und Schulbildung kontrolliert. Es gibt es auch keine signifikanten Interaktionen zwischen Bildung und Staatsangehörigkeit, so dass nicht angenommen werden kann, dass gerade minder qualifizierte Ausländer wegen größerer Sprachprobleme dazu neigen, den vertraulichen Fragebogen nicht auszufüllen.

Es liegt auch kein signifikanter linearer Alterseffekt vor. Selbst die Modellierung eines kurvilinearen Einflusses von Alter, gemessen über das quadrierte Lebensalter der Befragten, erbrachte keinen Erkenntnisgewinn. Die theoretisch erwarteten Erinnerungsprobleme für frühere Delikte oder Tendenzen zum sozial erwünschten Antwortverhalten bei Fragen zu zukünftigen Delikten liegen bei den älteren Befragten nicht in der Weise oder in dem Ausmaß vor, dass es zu systematischen Antwortverweigerungen kommt.

Der *Einfluss subkultureller Normen* auf die interessierende Antwortreaktion wird über die Verhaltensbeurteilung für vier Delikte (Ladendiebstahl, Steuerhinterziehung, Schwarzfahren und Trunkenheit am Steuer), die im anonymen Fragenbogen selbst Gegenstand sind, gemessen.⁶ Bei Befragten, die das „Schwarzfahren“ uneingeschränkt akzeptieren, ist die Antwortverweigerung sehr wahrscheinlich (*Modell 1* bzw. *3*). Hingegen neigen Befragte, welche Steuerhinterziehung als moralisch unbedenklich ansehen, eher dazu, den anonymen Fragenbogen auszufüllen als Befragte, welche den Steuerbetrug missbilligen. Das ist allerdings nur dann der Fall, wenn zusätzlich Interviewer- und Interviewereffekte kontrolliert werden (siehe *Modell 3*). Steuerhinterziehung wird wegen der damit verbundenen Einkommensvorteile eher von Befragten mit höherer Bildung und aus höheren Sozialschichten als moralisch unbedenklich angesehen, und das sind gleichzeitig Sozialgruppen, die generell antwortbereit sind (vgl. Mehlkop und Becker 2003). Moralische Beurteilungen von Massendelikten ohne direkte Opfer – etwa das Schwarzfahren oder Trunkenheit am Steuer – wirken sich hingegen nicht auf das Antwortverhalten aus.

6 Der Wortlaut für diese Frage lautet: „Ich werde Ihnen gleich einzelne Karten überreichen, auf denen verschiedene Verhaltensweisen beschrieben sind. Bitte sagen Sie mir jeweils mit Hilfe dieser Liste, ob Sie persönlich das beschriebene Verhalten für sehr schlimm, ziemlich schlimm, weniger schlimm oder für überhaupt nicht schlimm halten.“

Tabelle 1: Determinanten für das Ausfüllen des anonymen Fragebogens (,sealed envelope technique') – Logistische Regression (odds ratio)

	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5
<i>Befragtenmerkmale</i>					
Deutsche Staatsangehörigkeit	2,16**		2,07*		1,97***
Mittlere Reife	1,14		1,07		1,11
Fachhochschulreife oder Abitur	1,42		1,42		1,41
Lebensalter in Jahren	1,00		1,00		1,00
<i>Hohe Akzeptanz von ...</i>					
Ladendiebstahl	0,64*		0,61*		
Steuerhinterziehung	2,13*		2,15*		
Schwarzfahren	0,60		0,60		
Trunkenheit am Steuer	1,22		1,24		
<i>Interview- und Interviewereffekte</i>					
Anwesenheit haushaltsfremder Personen		0,29*	0,29*		0,32*
Männlicher Interviewer		1,026	1,025		1,031
Lebensalter der Interviewer in Jahren		0,98*	0,97**		0,98*
Erfahrung als Interviewer: bis 2 Jahre		0,78	0,67		0,76
Erfahrung als Interviewer: 2 bis 5 Jahre		0,68*	0,62*		0,67*
Erfahrung als Interviewer: 5 bis 10 Jahre		2,92***	3,16***		2,80***
<i>Beurteilung des Interviewverlaufs</i>					
Geringe Interviewbereitschaft				0,74	
Hohe Antwortbereitschaft				1,63**	
Korrekturhäufigkeit beim Interview				-0,92**	
Pseudo-R ² (Cox & Snell)	0,015	0,027	0,044	0,014	0,034
Pseudo-R ² (Nagelkerke)	0,029	0,050	0,083	0,026	0,062
N (Verweigerung)	179	214	179	214	210
N (Befragte)	1.437	1.605	1.437	1.613	1.587

* $p \leq 0,05$; ** $p \leq 0,01$; *** $p \leq 0,001$

Datenbasis: ALLBUS 2000 – eigene Berechnungen

Die *Interview- und Interviewereffekte*, die aufgrund der verfügbaren Datenlage berücksichtigt werden konnten, wirken sich hauptsächlich in der theoretisch erwarteten Weise auf das Antwortverhalten aus. Sind *Personen beim Interview anwesend, die nicht dem Haushalt der Befragten angehören*, treten eher Antwortverweigerungen auf, als wenn Ehe- bzw. Lebenspartner, andere Familienangehörige oder keine dritte Personen anwesend sind (*Modell 2 und 3*).

Das (männliche) *Geschlecht des Interviewers* hat entsprechend den theoretischen Annahmen zur Wirksamkeit der SET keinen signifikanten Einfluss auf das Ausfüllen des vertraulichen Fragebogens. Bei früheren Datenanalysen wurde zusätzlich die Geschlechterkonstellation der Interviewer und Befragten berücksichtigt. Selbst dann sind keine Auswirkungen dieser sichtbaren Interviewereigenschaften auf das Antwortverhalten der Befragten festzustellen. Hingegen sinkt wie erwartet mit zunehmendem *Alter des Interviewers* die Bereitschaft der Befragten trotz zugesicherter Anonymität, Fragen zu ihrer eigenen Delinquenz zu beantworten.

Schließlich wirken sich *Erfahrungen als Interviewer* in unterschiedlicher Weise auf das Antwortverhalten der Befragten aus. Die Erfahrung als Interviewer wird über die Dauer gemessen, mit der die Interviewer bei dem Institut beschäftigt sind, welches mit der Datenerhebung für den ALLBUS 2000 beauftragt wurde. Aus methodischer Sicht ist zu berücksichtigen, dass diese Variable für einige Interviewer nur Näherungswerte darstellen dürften, weil die Beschäftigung als Interviewer bei anderen Instituten mangels Daten nicht kontrolliert werden kann. Offenbar treten bei den Interviewern mit einer Beschäftigungsdauer von 2 bis 5 Jahren bei ein und demselben Erhebungsinstitut eher Antwortverweigerungen auf als bei „alten Hasen“, die mehr als 10 Jahre für dieses Institut tätig sind, während sich Interviewer mit einer Erfahrung von weniger als zwei Jahren nicht mehr oder weniger Antwortverweigerungen registrieren mussten als Interviewer mit einer Erfahrung von mehr als 10 Jahren. Am ehesten erfolgt das Ausfüllen des vertraulichen Fragebogens bei den Interviewern, die zwischen 5 und 10 Jahren für das Erhebungsinstitut tätig sind. Es gibt also *keinen linearen, aber einen kurvilinearen Effekt der Erfahrung als Interviewer* auf die Beantwortung des vertraulichen Fragebogens.

Betrachten wir abschließend die *subjektiven Einschätzungen des Interviewverlaufs durch die Interviewer*, die ausschließlich dem Interviewprotokoll entnommen wurden. Diese subjektiven Einschätzungen stehen in unserem Fall stellvertretend für eine *Vielzahl unbeobachteter Prozesse im Interview*. Es erhärten sich sowohl weitere theoretische Annahmen auch bereits dargestellte empirische Befunde (*Modell 4*). Zunächst hängt die *allgemeine Interviewbereitschaft* nicht mit der Neigung der Befragten zusammen, den anonymen Fragebogen auszufüllen. Sind abgesehen von den individuellen Motivationen und Fähig-

keiten der Befragten die methodischen Voraussetzungen erfüllt, um die generelle *Antwortbereitschaft* der Befragten zu fördern, dann werden auch zumeist die Fragen zum eigenen delinquenten Handeln beantwortet. Nicht geklärt werden kann jedoch, ob auch das Ausfüllen des vertraulichen Fragebogens in die Bewertung der Antwortbereitschaft mit eingeflossen ist. Sollte dies der Fall sein, dann dürfte deren Anteil an der gesamten Bewertung des Antwortverhaltens angesichts der großen Zahl von Fragen zu anderen Sachverhalten geringfügig sein. Letztlich gibt es Hinweise dafür, dass Bedingungen, die zu indifferenten und inkonsistenten Befragungssituationen führen, mit verantwortlich dafür sind, dass der vertrauliche Fragebogen nicht ausgefüllt wird. So gehen *häufige Korrekturen der Antworten während des Interviews* mit der Weigerung einher, Fragen zum eigenen delinquenten Verhalten zu beantworten, und je häufiger Revisionen von Antworten vorgenommen werden, desto wahrscheinlicher ist es, dass der vertrauliche Fragebogen unausgefüllt bleibt.⁷

Konsequenzen des selektiven Antwortverhaltens für Modellschätzungen

Die vorgelegten Analysen belegen, dass die Antwortverweigerungen *nicht* zufällig sind, sondern selektiv in Bezug auf individuelle Merkmale der Befragten und Interviewer erfolgen sowie auf Intervieweffekte zurückzuführen sind. Nun wissen wir nicht, wie die Verweigerer geantwortet hätten, hätten sie den vertraulichen Fragebogen ausgefüllt. Daher untersuchen wir, ob sich diese systematischen Antwortverweigerungen bei multivariaten Analysen des delinquenten Handelns der Befragten zu verzerrten Schätzergebnissen führen.

Im Folgenden beschränken wir uns auf die im vertraulichen Fragebogen angesprochenen Straftaten wie *Steuerhinterziehung, Ladendiebstahl, Schwarzfahren und Trunkenheit am Steuer* und betrachten ausschließlich die Resultate für die standardisierte Inklusionswahrscheinlichkeit (*Tabelle 2*).⁸ Ihre Operationalisierung basiert auf dem *Modell 5 in Tabelle 1*.

7 Zusätzliche, hier aus Platzgründen nicht dokumentierte Analysen haben beispielsweise ergeben, dass die Interaktion von Ausländerstatus und Korrekturhäufigkeit nicht signifikant zur Aufklärung dieses Sachverhaltes beitragen. Verständnis der deutschen Sprache oder Sprachschwierigkeiten gehen demnach nicht zwingend mit der Korrekturhäufigkeit von Antworten einher.

8 Der subjektiv erwartete Nutzen einer strafbaren Handlung B konnte nur indirekt über die Verhaltensbeurteilung für einzelne Straftaten abgebildet werden, weil im ALLBUS eine direkte Messung eines erwarteten Gewinns durch eine Straftat nicht vorgenommen wurde. Mit dieser moralischen Bewertung des delinquenten Verhaltens „kann festgestellt werden, ob eine bestimmte Verhaltensnorm überhaupt in der Bevölkerung bzw. in einzelnen Subpopulationen Geltung besitzt, anders ausgedrückt: ob eine bestimmte Verhaltensweise überhaupt als abweichend einge-

Tabelle 2: Determinanten des kriminellen Handelns unter besonderer Berücksichtigung eines ‚selectivity bias‘ – Logistische Regression (odds ratio)

	Steuerhinterziehung		Ladendiebstahl		Schwarzfahren		Trunkenheit	
	1	2	1	2	1	2	1	2
Nutzen <i>B</i>	3,39***	3,40***	2,72***	2,70**	4,16***	3,97***	5,11***	5,35***
Risiko <i>p</i>	0,98***	0,98***	0,99*	0,99*	0,98***	0,99***	0,98***	0,98***
Strafe <i>C</i>	0,43***	0,73***	0,66	0,69	0,37***	0,37***	0,61*	0,62*
Inklusion λ		1,25**		0,96		1,37***		1,00
Pseudo-R ²	0.084	0.089	0.011	0.010	0.083	0.093	0.088	0.076
N	1.187	1.187	1.251	1.251	1.269	1.269	1.222	1.222

* $p \leq 0,05$; ** $p \leq 0,01$; *** $p \leq 0,001$

1 ohne Kontrolle der Selektivität

2 mit Kontrolle der Selektivität

Datenbasis: ALLBUS 2000 – eigene Berechnungen

Für die beiden Delikte „Ladendiebstahl“ und „Trunkenheit am Steuer“ sind die Inklusionswahrscheinlichkeiten insignifikant. Offensichtlich hätten die Modellschätzungen für diese Delikte keiner Korrektur wegen einer Stichprobenverzerrung durch Antwortverweigerungen bedurft.

Jedoch liegen bei den Delikten wie Steuerhinterziehung oder Schwarzfahren *positive Selektivitätsverzerrungen* vor, wonach Personen, die den vertraulichen Fragebogen ausgefüllt haben, auch eher dazu neigen, falsche Angaben bei der Einkommensteuererklärung

stift wird“ (Wasmer et al. 1991: 25). Man kann durchaus annehmen, dass diese Beurteilung, die von „sehr schlimm“ bis „gar nicht schlimm“ reicht, mit dem erwarteten Nutzen dieser Straftat zusammenhängt. Ein Verstoß gegen die selbstverpflichtete Gesetzestreue als internalisierte Norm versucht Kosten. Der Kostenfaktor wird mit einer Dummy-Variablen und basiert auf der Frage zu der Meinung, ob man sich an Gesetze unbedingt immer zu halten habe. Die Referenzkategorie sind Personen, die dieser Meinung nicht zustimmen. Sicherlich deckt dieser Indikator für subjektiv erwarteten Kosten von Straftaten C nur einen Teil der erwarteten Gesamtkosten ab. Mangels Informationen beim ALLBUS 2000 gibt es keine theoretisch sinnvolle Alternative zu diesem Indikator (vgl. Mehlkop und Becker 2004). Im ALLBUS 2000 sind neben erwarteten Gewinnen und Kosten einer Straftat auch die subjektiven Entdeckungswahrscheinlichkeiten *p* – wie von Opp (1989: 426) empfohlen – direkt oder indirekt durch Interviews gemessen worden (vgl. Paternoster et al. 1982: 1255). Die Befragten sollten auf einer fünfstufigen Skala von „sehr wahrscheinlich“ bis „sehr unwahrscheinlich“ angeben, wie wahrscheinlich es ist, dass man sie beim Begehen der strafbaren Handlung erwischt. Für die multivariaten Analysen wurden die nominalen Angaben in prozentuale Wahrscheinlichkeitsangaben – abgestuft von 0 Prozent für „sehr unwahrscheinlich“ bis 100 Prozent für „sehr wahrscheinlich“ – transformiert.

oder beim Lohnsteuerjahresausgleich zu machen, um weniger Steuern zahlen zu müssen, oder öffentliche Verkehrsmittel zu benutzen, ohne einen gültigen Fahrausweis zu besitzen. Demnach sind in der Analysestichprobe diejenigen Befragten überrepräsentiert, die in Zukunft (wieder) Steuern hinterziehen würden oder (wiederum) zum Schwarzfahren bereit wären.

5. Diskussion

Ziel des vorliegenden Beitrages war es, die Wirksamkeit der ‚sealed envelope technique‘ (SET) zu untersuchen. Die SET wird als ein geeignetes Vorgehen gesehen, Befragten absolute Anonymität und Vertraulichkeit bei heiklen wie unangenehmen Fragen zu deren Delinquenz zu garantieren. Dadurch sollten sowohl verzerrte Antworten wie Antwortverweigerungen unwahrscheinlicher werden. Es wurde zunächst der Frage nachgegangen, warum bestimmte Personengruppen trotz zugesicherter Anonymität, das Selbstaussfüllen eines vertraulichen Fragebogens zum delinquenten Verhalten verweigern, und welche Bedingungen des Interviews Tendenzen zur Antwortverweigerung fördern. Um theoretisch fundierte Antworten und empirisch überprüfbare Hypothesen abzuleiten, wurde das werterwartungstheoretische Modell des Befragtenverhaltens von Esser (1986) herangezogen. Des Weiteren wurden die *Konsequenzen selektiver Antwortreaktionen* für die Datenauswertung untersucht, die für statistische Analysen des delinquenten Handelns entstehen können. Als empirisches Beispiel wurde die ökonomische Theorie der Kriminalität von Gary S. Becker (1968) gewählt. ‚Selectivity bias‘ sollten dabei mit dem Korrekturverfahren von Heckman (1979) aufgefunden gemacht werden.

Die empirischen Analysen erfolgten auf Basis der Daten des *ALLBUS 2000*. Mangels Informationen konnten keine direkten Messungen von subjektiven Bewertungen und Erwartungen der Befragten vorgenommen werden. In dieser Hinsicht war die empirische Anwendung des Erklärungsmodells von Esser (1974, 1986) unvollständig. Dieses Faktum ist unbefriedigend, aber bei Sekundäranalysen zumeist die Regel. Da helfen in Zukunft nur direkte Messung der Präferenzen, Bewertungen und Erwartungen des Befragten sowie umfassende Dokumentation des Interviewverlaufs im experimentellen Design.

Neben individuellen Merkmalen der Befragten – etwa ihre Kompetenzen in deutscher Sprache oder ihre Einstellungen zu Straftaten – beeinflussen auch bei Verwendung eines vertraulichen, selbst auszufüllenden Fragebogens die Merkmale der Interviewer und der Interviewersituation die Tendenzen des Befragten für Antwortverweigerungen oder für „wahre“ Antworten zum delinquenten Handeln. Vor allem die Anwesenheit dritter, nicht zum Haushalt der Befragten zugehörigen Personen, das Alter der Interviewer und ihre Erfahrung als Interviewer konterkariert das Bestreben, über den vertraulichen Fragebogen

die Anonymität der Angaben zum delinquenten Handeln zu garantieren. Insgesamt dominieren die Interview- und Interviewereffekte über die individuellen Effekte der Befragten und sind am ehesten für das selektive Antwortverhalten bei Fragen zum delinquenten Verhalten der Befragten verantwortlich. Es ist allerdings nicht ausgeschlossen, und dies kann mit den Daten des ALLBUS nicht empirisch überprüft werden, dass möglicherweise eher *unbeobachtete Interview- und Interviewereffekte* (etwa Art und Weise der Erläuterung der SET durch den Interviewer, Auftreten und Gebaren des Interviewers während dieser Prozedur, Abweichung von den Anweisungen für die Durchführung der ‚sealed envelope technique‘ oder Verlautbarung von Einstellungen und Erwartungen seitens des Interviewers in Bezug auf die Delinquenz) ausschlaggebend für verweigerte Antworten sind.

Das auch bei Einsatz der SET selektive Antwortverhalten führte in zwei Fällen zu verzerrten Schätzergebnissen. So waren Befragte überrepräsentiert, die – bei Kontrolle der erwarteten Vorteile ihres strafbaren Handelns, der erwarteten Kosten bei Normverletzung und der subjektiven Sanktionswahrscheinlichkeit – berichteten, dass sie in der Zukunft (weiterhin) Steuern hinterziehen oder „Schwarzfahren“ würden. Diese Befunde weisen darauf hin, dass entgegen den Alltagsvermutungen die Antwortverweigerer seltener Straftaten begehen würden als die Gruppe der Befragten, die Fragen zu ihrer früheren und beabsichtigten Delinquenz beantworten. Zur Klärung der Frage, wie dieses Faktum zustande kommt, bedarf es weiterer inhaltlicher und methodisch orientierter Untersuchungen.

In Zukunft sind generell bei Beibehaltung der erfolgreichen ‚sealed envelope technique‘ verstärkte Anstrengungen vorzunehmen, bei heiklen, unangenehmen oder bedrohlichen Fragen die Interviewsituation zu „neutralisieren“. Gerade in Befragungssituationen, die durch inkonsistentes Befragtenverhalten gekennzeichnet sind, müssen Interview- und Interviewereffekte gezielt ausgeschaltet werden. Insbesondere könnte dann die *postalische Befragung (mail survey)* die geeignetere Befragungsmethode als das ‚face-to-face-Interview‘ sein (Sudman und Bradburn 1982). Denn bei postalischen Befragungen sind Anonymitätszusicherungen eindeutig und Interviewereffekte ausgeschlossen. Ebenso könnten die ungünstigen Auswirkungen auf das Antwortverhalten, die mit der Anwesenheit dritter Personen verbunden sind, minimiert werden, da es recht unwahrscheinlich sein dürfte, dass nicht zum Haushalt der Befragten zugehörigen Personen beim Beantworten des postalisch zugesandten Fragebogens anwesend sein werden.

Trotz aller methodischen Anstrengungen wird es weiterhin ‚item nonresponse‘ bei Fragen zum delinquenten Handeln geben, die in der Motivation der Befragten selbst begründet ist (vgl. Schnell 1997). Jedoch gilt es zum einen systematische Ausfälle zu vermeiden und zum anderen, falls diese sich nicht vermeiden lassen, statistische Verfahren weiter zu entwickeln, um daraus resultierende Verzerrungen bei den statistischen Auswertungen

auszugleichen (siehe Pötter und Rendtel 1993). Dazu ist aber notwendig, mehr Informationen über den Verlauf des Interviews sowie über den Prozess der Antwortverweigerungen einschließlich ihrer Gründe zu sammeln. Ansonsten bleiben mangels theoretischer Fundierung die statistischen Verfahren für den Ausgleich von ‚selectivity bias‘ beliebig (vgl. Heckman 1979).

Literatur

Becker, G.S., 1968: Crime and Punishment: An Economic Approach, *Journal of Political Economy* 76: 169-217.

Becker, R., 2001: Reliabilität von retrospektiven Berufsverlaufsdaten. Ein Vergleich zwischen dem öffentlichen Dienst und der Privatwirtschaft anhand von Paneldaten, *ZUMA-Nachrichten* 25: 29-56.

De Leeuw, E.D., 2001: Reducing Missing Data in Surveys: An Overview of Methods, *Quality & Quantity* 35: 147-160.

Diekmann, A., 1980: Die Befolgung von Gesetzen. Empirische Untersuchungen zu einer rechtssoziologischen Theorie, Berlin: Duncker & Humblot.

Diekmann, A., 1995: Empirische Sozialforschung. Reinbek: Rowohlt.

Esser, H., 1974: Der Befragte. S. 107-145 in: Jürgen van Koolwijk u.a.: Erhebungsmethoden: Die Befragung. Techniken der empirischen Sozialforschung (4. Band), München: Oldenbourg.

Esser, H., 1986: Können Befragte lügen? Zum Konzept des „wahren Wertes“ im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung, *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 38: 314-336.

Heckman, J.J., 1979: Sample Selection Bias As a Specification Error. *Econometrica* 47: 153-161.

Hedström, P./Swedberg, R., 1998: Social mechanisms: An introductory essay. S. 1-31 in: Hedström, P./Swedberg, R. (Hg.): Social Mechanisms. An Analytical Approach to Social Theory. Cambridge: University Press.

Hindelang, M. J./Hirschi, T./Weis, J.G., 1979: Correlates of Delinquency: The Illusion of Discrepancy between Self-Report and Official Measures, *American Sociological Review* 44: 995-1014.

Koch, A./Wasmer, M./Harkness, J./Scholz, E., 2001: Konzeption und Durchführung der "Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS) 2000. ZUMA-Methodenbericht 2001/05. Mannheim: ZUMA.

Mehlkop, G./Becker, R., 2003: Soziale Schichtung und Delinquenz. Eine empirische Anwendung eines Rational Choice-Ansatzes mit Hilfe von Querschnittsdaten des ALLBUS 1990 und 2000 (erscheint in: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 56: 122-153).

Opp, K.-D., 1989: The Economics of Crime and the Sociology of Deviant Behaviour. A Theoretical Confrontation of Basic Propositions, *Kyklos* 42: 405-430.

Paternoster, R./Saltzman, L.E./Chiricos, Th.G./Waldo, G.P., 1982: Perceived Risk and Deterrence: Methodological Artifacts in Perceptual Deterrence Research, *The Journal of Criminal Law & Criminology* 73: 1238-1258.

Pötter, U./Rendtel, U., 1993: Über Sinn und Unsinn von Repräsentativitätsstudien, *Allgemeines Statistisches Archiv* 77: 269-280.

Schnell, R., 1997: Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen, Opladen: Leske+Budrich.

Schwarz, N./Hippler, H.-J. /Noelle-Neumann, E., 1994: Retrospective Reports: The Impact of Response Formats. S. 187-199 in: Schwarz, N./Sudman, S. (Hg.): *Autobiographical Memory and the Validity of Retrospective Reports*. New York: Springer.

Sudman, S./Bradburn, N.M., 1974: *Response Effects in Surveys: A Review and Synthesis*, Chicago: Aldine.

Sudman, S./Bradburn, N.M., 1982: *Asking Questions*. San Francisco: Jossey-Bass Publishers.

Sudman, S./Bradburn, N.M./Schwarz, N., 1996: Thinking about Answers. The Application of Cognitive Processes to Survey Methodology, San Francisco: Jossey-Bass Publishers.

Urban, D. 1993: Logit-Analyse. Statistische Verfahren zur Analyse von Modellen mit qualitativen Response-Variablen, Stuttgart: Fischer.

Wasmer, M./Koch/Wiedenbeck, M., 1991: Methodenbericht zur „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS) 1990, ZUMA-Arbeitsbericht 91/13.

Korrespondenzadressen

*Prof. Dr. Rolf Becker
Universität Bern
Institut für Pädagogik und Schulpädagogik
Abteilung Bildungssoziologie
Muesmattstraße 27
CH-3012 Bern
Telefon: +41(0)31 631 5351
Telefax: +41(0)31 631 5352
email: Rolf.Becker@sis.unibe.ch*

*Dipl. Soz. Ralph Günther
Technische Universität Chemnitz
Philosophische Fakultät
Institut für Soziologie
Professur für Empirische Sozialforschung
Reichenhainer Strasse 41
D-09107 Chemnitz
Telefon: +49(0)371 531-3925
Telefax: +49(0)371 531-4450
email: Ralph.Guenther@phil.tu-chemnitz.de*

KINDER UND JUGENDLICHE ALS BEFRAGTE

Feldexperimente zum Antwortverhalten Minderjähriger

MAREK FUCHS

Die Umfragemethodologie ist hauptsächlich für Studien in Erwachsenenpopulationen entwickelt worden. In den letzten Jahren werden jedoch zunehmend Befragungen durchgeführt, bei denen die Verhaltensweisen und Sichtweisen von Kindern und Jugendlichen direkt von ihnen ermittelt werden. Die Befragung von Eltern oder anderen Stellvertretern verliert demgegenüber an Bedeutung. Nimmt man an, dass die kognitiven Fähigkeiten von Kindern und Jugendlichen im Vergleich zu denen von Erwachsenen noch nicht voll entwickelt sind, dann stellt sich die Frage, ob Kinder und Jugendliche die komplizierten Aufgaben bei der Beantwortung eines Fragebogens bewältigen können.

Der vorliegende Beitrag geht dieser Frage nach, indem die Qualität der von Kindern und Jugendlichen erhaltenen Daten nach Alter und kognitiven Fähigkeiten verglichen wird. Dabei konzentrierten wir uns auf Antwortfehler und schließen von der Größe der Antwortfehler auf die vorgängigen kognitiven Prozesse. Ausgehend von einem Überblick über den Stand der Forschung werden die Ergebnisse von 7 Feldexperimenten zu Antwortfehlern berichtet, die in drei größer angelegte Befragungen von Kindern und Jugendlichen integriert waren ($n=4.206$, $n=5.042$, $n=820$): Experimente zur Fragereihenfolge, zur Antwortreihenfolge, zu Skalen-Effekten und zu den Effekten der numerischen Werte an den Antwortalternativen. Die Ergebnisse zeigen, dass Kinder und Jugendliche Fragebogenfragen von einem kognitiv weniger entwickelten Stand aus beantworten: Sie bearbeiten jede einzelne Fragebogenfrage stärker segmentiert und isoliert, zugleich lassen sie sich bei der Beantwortung der jeweiligen Frage stärker von den darin enthaltenen Informationen leiten.

Survey methodology is predominantly developed for studies in adult populations. In recent years, however, many surveys have addressed samples of children and juveniles. Today, data on children's behavior, attitudes and beliefs are usually collected directly from them. Proxy-reporting from parents or other adults taking care of the children is no longer seen as an adequate and appropriate means of data collection. At the same time, given that their cognitive skills may still be developing, one may wonder whether

children can adequately perform the difficult tasks associated with responding to a questionnaire. As a result the quality of data obtained from children might be worse than that collected from adult populations.

In this paper we assess the quality of data obtained from children and juveniles. We focus on response error and compute the sizes of response error for specific age groups and groups with different levels of cognitive functioning. Against the background of an extensive literature review, we analyze field experiments that were part of three large-scale surveys among children and juveniles aged 10 to 21 ($n=4.206$, $n=5.042$, $n=820$). The results of 7 experiments on response order, question order, scale effects and the effects of the numeric values associated to the response categories are reported. The results indicate that younger and less well-educated children answer survey questions from a cognitively less sophisticated perspective than do adults. Children and juveniles process survey questions in a more segmented fashion and view each as separate entities. In addition, in answering questions they are influenced to a greater degree than adults by the information provided in the question text and by the response categories of each given question.

1. Einleitung und Fragestellung

Die Theorie des Frage-Antwort-Prozesses ist relativ gut entwickelt, ebenso die Regeln für die Konstruktion von „guten“ Fragebögen. Als Ergebnis dieser methodologischen Bemühungen kann man heute auf eine Reihe von Überblicksdarstellungen zur Datenqualität (Groves 1989; Lyberg et al. 1997) und zur Fragebogenkonstruktion (Schuman/Presser 1981; Sudman et al. 1996; Tourangeau et al. 2000) zurückgreifen. Diese Darstellungen beziehen sich im Wesentlichen auf Umfragen unter Erwachsenen; in den letzten Jahrzehnten wurde jedoch eine Reihe von Studien mit Kindern und Jugendlichen durchgeführt, bei denen deren Verhaltensweisen, deren Einstellungen und sonstige Informationen von ihnen selbst erhoben, und nicht auf Eltern oder anderen Stellvertretern (Proxy) zurückgegriffen wurde (in Deutschland z.B. das DJI-Kinderpanel, aber auch Studien zur Devianz von Kindern und Jugendlichen oder kommerzielle Untersuchungen). In der Literatur (Holiday/Turner-Henson 1989; Scott 1997; Borgers et al. 2000) wird in diesem Zusammenhang darauf hingewiesen, dass Umfragen unter Kindern und Jugendlichen deren kognitiven und sozialen Entwicklungsstand zu berücksichtigen haben.¹ Denn die begrenzten kognitiven Ressourcen der kindlichen Befragten könnten u. U. zu einem eingeschränkten Fragever-

1 Eine sehr nette Verulkung der Methodenforschung zur Befragung von kleinen Kindern findet sich bei Miller 1976.

ständnis und zu anderen Beeinträchtigungen der Datenqualität führen, weil die jungen Befragten nicht oder nur eingeschränkt über die Voraussetzungen zur Beantwortung eines Fragebogens verfügen.

Ausgehend von der Literatur zur Psychologie des Umfrageinterviews (z.B. Sudman et al. 1996) stellt sich der Frage-Antwort-Prozess als komplexe Interaktion des Befragten mit dem Fragebogen und anderen Elementen der Umfrage dar. In dieser Perspektive gliedert sich die Aufgabe des Befragten bei der Beantwortung eines einzelnen Items in verschiedenen Phasen: Zunächst muss der Frageinhalt verstanden werden bevor nach relevanten Informationen zu ihrer Beantwortung gesucht werden kann. Darauf aufbauend wird die eigentliche Antwort generiert, bevor diese hinsichtlich des angebotenen Antwortschemas formatiert und schließlich unter Umständen editiert wird. Bei der Abarbeitung dieser Phasen des Frage-Antwort-Prozesses berücksichtigt der Befragte die Frageformulierung und die dargebotenen Antwortkategorien der aktuellen Frage und ebenso die vorausgehenden – und zum Teil auch die nachfolgenden – Fragen als signifikanten Kontext (für eine umfassende Übersicht vgl. auch Tourangeau et al. 2000). In Anbetracht der Komplexität dieses Vorgangs stellt sich die Frage „whether children can adequately perform these difficult tasks“ (de Leeuw/Otter 1995: 251) – denn Kinder verfügen im Vergleich zu Erwachsenen über weniger entwickelte kognitive Fähigkeiten.

Die kognitive Entwicklung von Kindern und Jugendlichen wird in der Regel als Stufenfolge konzeptualisiert (vgl. u. a. Dockrell et al. 2000): Folgt man den theoretischen Ansätzen von Piaget (1948) oder Kohlberg (1976), so ist die menschliche Entwicklung nicht als ein diskreter Schritt von Kindheit zum Erwachsenenstatus zu verstehen. Vielmehr verläuft die Entwicklung von Kindern und Jugendlichen als kontinuierlicher Prozess über einen längeren Zeitraum, in dem nacheinander verschiedene konsekutive Stufen der kognitiven und sozialen Entwicklung erreicht werden. Daher können Kinder und Jugendliche nicht als homogene Gruppe verstanden werden, sondern es muss zwischen verschiedenen Stufen der kognitiven Kapazität differenziert werden. Ausgehend von diesen Theorien wird in der Literatur zur Umfrageforschung mit Minderjährigen die Vermutung vertreten, dass Kinder ab einem Alter von etwa 10 Jahren in der Lage sind, standardisierte Befragungen zu bearbeiten (vgl. etwa Scott 1997 und Borgers et al. 2000 für Überblicksdarstellungen; vgl. Benson/Hocevar 1985 für Aussagen zu Grundschulkindern). Es wird angenommen, dass Kinder dieses Alters bereits eine spezifische Stufe der formalen Operation (Piaget 1948) oder – in Kohlbergs (1976) Terminologie – die erste Stufe des konventionellen Levels der moralischen Entwicklung erreicht haben. Daher sind sie in der Lage und willens, abstrakte Inhalte zu verarbeiten und Regeln zu folgen, die unabhängig von spezifischen Subjekten in konkreten Interaktionssituationen Geltung beanspruchen. Es ist jedoch offensichtlich, dass 10-jährige Kinder Fragebogenfragen vor dem Hintergrund weniger

entwickelter kognitiver Fähigkeiten betrachten und bearbeiten als 17-jährige Jugendliche oder Erwachsene, und dass sich daher ihr Antwortverhalten in standardisierten Befragungen unterscheidet. Es stellt sich damit die Frage, wie sich Kinder, Jugendliche und junge Erwachsene bei der Bearbeitung eines standardisierten Fragebogens unterscheiden.

Bisher liegen jedoch nur wenige empirische Analysen zum Antwortverhalten von Kindern und jüngeren Jugendlichen in standardisierten Befragungen vor; außerdem handelt es sich dabei im Wesentlichen sekundäranalytische Studien auf der Basis existierender Datensätze. Allgemein wird jedoch davon ausgegangen, dass die begrenzte kognitive Kapazität jüngerer Befragter die Datenqualität negativ beeinflusst: Basierend auf einem dreiwöchigen Panel zur politischen Sozialisation kann Villancourt (1973) zeigen, dass die Antwortstabilität von Kindern – wie bei den Antworten von Erwachsenen – erheblich variiert. Allerdings wurde kein Vergleich mit Jugendlichen oder Erwachsenen durchgeführt. Marsh (1986) stellt bei einem Vergleich verschiedener Altersgruppen fest, dass „younger children and children with poorer reading skills are less able to response appropriately to negative items and that this effect biases the interpretation of their responses“ (1986: 37; vgl. auch Hershey/Hill 1976).

Amato und Ochiltree (1987) zeigen, dass die Datenqualität einer Gruppe von 8- und 9-Jährigen signifikant niedriger ist, als die einer Vergleichsgruppe von 15- und 16-Jährigen. In einer Studie von Borgers und Kollegen (2000) erwies sich die Anzahl der absolvierten Schuljahre als aussagekräftiger Prädiktor für die interne Konsistenz von Multi-Item-Skalen. Dies unterstützt die Hypothese, wonach die Datenqualität mit zunehmender kognitiver Kapazität ansteigt. Hinzu kommt, dass schlechte Lesekompetenzen einen negativen Einfluss auf die Datenqualität haben (Borgers et al. 2000). Weiterhin konnte gezeigt werden, dass mehrdimensionale Antwortskalen für jüngere Kinder mit größeren Schwierigkeiten verbunden waren und einen höheren Item-Nonresponse produzierten als bei älteren Kindern. Folgt man Borgers und Hox (2001) ist dies auf deren kognitives Entwicklungsstadium zurückzuführen.

Unterstützend können hier die Ergebnisse von de Leeuw und Otter (1995) herangezogen werden, die eine klare Interaktion zwischen dem Alter und dem Effekt unklarer Fragen nachweisen konnten. Danach war der negative Effekt mehrdeutiger Fragen auf die Datenqualität bei jüngeren Kindern stärker ausgeprägt als bei älteren Kindern und Jugendlichen, die mit der Ambiguität besser umgehen konnten. Neben dem Alterseffekt fanden Borgers und Hox (2000) ebenfalls einen Einfluss des Leistungsstandes in der Schule. Fasst man diese vorliegenden Ergebnisse zusammen, so stützen sie die Hypothese, wonach das geringe Alter von Kindern und jüngeren Jugendlichen und zudem eine geringe schulische Leistungsfähigkeit (als Indikatoren für das kognitive Entwicklungsstadium) einen negati-

ven Einfluss auf die Bearbeitung von Fragebogenfragen und damit auf die Datenqualität haben.

Diese These erfährt durch Erkenntnisse zum Antwortverhalten von Erwachsenen Unterstützung: Folgt man Krosnick (1992), dann sind umfangreiche Gedächtnisprozesse und kognitive Aktivitäten für Befragte mit geringer kognitiver Kapazität weniger angenehm, und sie führen diese Aufgaben weniger gerne aus. Ohne diese Aktivitäten wird jedoch ein spezifisches Antwortverhalten wahrscheinlich, das Krosnick (1992) als „Satisficing“ bezeichnet: Anstatt den Frage-Antwort-Prozess vollständig und gründlich zu durchlaufen, unternehmen die dieses Verhalten zeigenden Befragten weniger Anstrengungen bei der Dechiffrierung einer Fragebogenfrage und bei der Suche nach einer Antwort. Entsprechend wird nur „der Form genüge getan“ und es wird z.B. die erste akzeptable Antwortalternative gewählt oder sonst ein „leichter“ Weg durch die Befragung gesucht.

Jedoch fehlt bisher eine experimentelle Überprüfung des Zusammenhangs von kognitiven Fähigkeiten und der Bearbeitung des Frage-Antwort-Prozesses durch Kinder – dazu versucht der vorliegende Beitrag beizutragen: In sieben größer angelegten Feldexperimenten wurden das Alter und das schulische Leistungsniveau als Indikatoren für die kognitive Kapazität benutzt (vgl. hierzu die Ausführungen im Abschnitt über die Methoden), und zu Antworteffekten, an denen sich die kognitiven Prozesse bei der Beantwortung einer Fragebogenfrage ablesen lassen, in Beziehung gesetzt. Dazu wird in einem ersten Schritt überprüft, ob diese Antworteffekte auch für Kinder und Jugendliche reproduziert werden können, bevor in einem zweiten Schritt der Einfluss von Alter und schulischer Leistungsfähigkeit auf die Größe dieser Effekte untersucht wird.

Ausgehend von der Vermutung, dass die kognitiven Fähigkeiten von Kindern weniger entwickelt sind als die von Jugendlichen bzw. Erwachsenen, lassen sich verschiedene Hypothesen über die Besonderheiten der Antworteffekte bei Kindern und Jugendlichen aufstellen, die im vorliegenden Beitrag getestet werden sollen. Diese Hypothesen folgen der generellen Annahme, dass Kinder und jüngere Jugendliche die einzelnen Fragebogenfragen stärker segmentiert und aus dem Fluss des Fragebogens herausgelöst interpretieren, weil ihnen die kognitiven Fähigkeiten fehlen, sich den Fragebogen bzw. einzelne Frageblöcke als thematische Einheit zu vergegenwärtigen und im Gedächtnis zu behalten. Zugleich lassen sie sich bei der Beantwortung jeder einzelnen Frage stärker von deren Merkmalen leiten, weil sie bei der Bewältigung des kognitiv anspruchsvollen Frage-Antwort-Prozesses auf die Unterstützung und Hilfe durch die in der Frage und in den Antwortkategorien enthaltenen Informationen angewiesen sind. Im Einzelnen werden folgende Hypothesen überprüft:

Fragereihenfolge: Befragte, die eine geringere kognitive Funktionalität aufweisen, können bei der Bearbeitung einer Frage vorausgehende Fragen und Informationen in geringerem Umfang berücksichtigen und daher sind bei ihnen Kontexteffekte – als Indikator für den Umfang der berücksichtigten vorausgehenden Information – weniger ausgeprägt. Entsprechend vermuten wir bei jüngeren Befragten und bei Befragten mit geringerer schulischer Leistungsfähigkeit kleinere Fragereihenfolgen-Effekte als bei älteren Befragten und Befragten mit besserer schulischer Leistungsfähigkeit.

Skalen-Effekte: Neben den vorausgehenden Fragen nutzen die Befragten auch Merkmale der aktuellen Frage und der vorgegebenen Antwortmöglichkeiten, um den Frageinhalt zu verstehen und um ihre Antworten zu generieren. So wird bei geschlossenen Häufigkeitsfragen neben dem Fragetext von den Befragten auch das vorgegebene Antwortkontinuum herangezogen, um eine Vorstellung von der Größenordnung der gesuchten Antwort zu bekommen – vor allem, wenn die gesuchte Antwort geschätzt werden muss und nicht aus der Erinnerung abgerufen werden kann. Entsprechend unserer generellen Annahme werden sich jüngere Befragte und solche mit schlechteren schulischen Leistungen bei der Beantwortung von geschlossenen Häufigkeitsfragen stärker von den vorgegebenen Antwortkategorien leiten lassen. Als Ergebnis werden wir bei ihnen größere Skalen-Effekte nachweisen können.

Numerische Werte: Neben den verbalisierten Antwortkategorien lassen sich die Befragten auch von den numerischen Werten an Antwortalternativen leiten. Von den einleitend diskutierten theoretischen Überlegungen ausgehend, lässt sich vermuten, dass sich Befragte mit geringer entwickelten kognitiven Fähigkeiten bei der Generierung des Frageverständnisses stärker an den Zahlenwerten, die den Antwortkategorien beigelegt sind, orientieren werden. Entsprechend sollte der Effekt unterschiedlicher numerischer Symbole bei jüngeren Befragten und solchen mit schlechteren schulischen Leistungen größer ausfallen.

Antwortreihenfolge: Weiter vermuten wir, dass die begrenzte kognitive Kapazität zu größeren Antwortreihenfolgen-Effekten führt. Kinder und Jugendliche mit einer geringer entwickelten kognitiven Kapazität sind weniger gut in der Lage, jeder Antwortkategorie die Aufmerksamkeit zuzuwenden und daher werden sie sich bei der Auswahl einer Antwortkategorie eher von deren sequenzieller Position leiten lassen. Dies führt bei jüngeren Befragten und solchen mit schlechteren schulischen Leistungen zu größeren Antwortreihenfolgen-Effekten (Recency- und Primacy-Effekte).

2. Methode

Diese Hypothesen werden anhand der Ergebnisse von Feldexperimenten überprüft, die auf drei größeren, selbst-administrierten Umfragen basieren,² die in den Jahren 1999 bis 2001 mit Kindern, Jugendlichen und jungen Erwachsenen im Alter von 10 bis 21 Jahren zu verschiedenen inhaltlichen Themen durchgeführt wurden ($N_1 = 4.206$; $N_2 = 5.042$; $N_3 = 820$):

Die *erste* Studie wurde im Frühjahr 1999 als zweite Welle einer laufenden Längsschnittuntersuchung zur „Gewalt in Schulen“ in Bayern durchgeführt (Fuchs et al. 2001). Eine disproportional geschichtete Klumpenstichprobe von bayerischen Schülern allgemein bildender und beruflicher Schulen wurde untersucht, um repräsentative Aussagen über die Gewalttätigkeit und die Opfererfahrungen der Schüler in Bayern machen zu können. Insgesamt wurden 4.206 Kinder und Jugendliche im Alter zwischen 10 und 21 Jahren (einige wenige waren älter) im Klassenzimmer mit einem selbst-administrierten Fragebogen befragt.

Die *zweite* Studie wurde im Frühjahr 2001 ebenfalls als selbst-administrierte Befragung zum Thema „Rechtsextremismus unter Jugendlichen“ durchgeführt (Fuchs et al. 2003). 5.042 Befragte im Alter zwischen 13 und 21 Jahren (einige wenige waren älter) bearbeiteten einen Fragebogen im Klassenzimmer-Setting. Auch hier haben wir eine disproportional geschichtete Klumpenstichprobe gezogen, um eine repräsentative Schülerpopulation der allgemeinbildenden und beruflichen Schulen in Bayern realisieren zu können.

Die *dritte* Studie beschäftigte sich mit Konflikten zwischen den Generationen. Die Studie wurde im Frühjahr 2000 im Bundesland Bayern als schriftlich-postalische Befragung durchgeführt, die auf einer Random-Registerstichprobe von 820 Befragten basierte, die nach Alter und Gemeindegröße geschichtet war. Die Besonderheit dieser Studie besteht darin, dass Jugendliche und junge Erwachsene zwischen 15 und 20 Jahren mit Erwachsenen sowie mit alten und hochbetagten Befragten verglichen werden können.

In diese Studien wurden Experimente zur Fragereihenfolge, zu Skalen-Effekten, zur Antwortreihenfolge und zu den numerischen Werten an Antwortvorgaben integriert. Die Befragten wurden zufällig zu zwei bzw. vier Fragebogen-Splits zugewiesen. Jede dieser Gruppen erhielt einen ansonsten identischen Fragebogen, der sich nur hinsichtlich der experimentell modifizierten Fragen unterschied. Durch den Vergleich der Antwortverteilungen der Splits können Antworteffekte identifiziert und berechnet werden. Aufgrund der beträchtlichen Fallzahlen besteht weiter die Möglichkeit, die ermittelten Antworteffekte

2 Zu Mode-Effekten bei Befragungen von Jugendlichen vgl. Naplaval/Oberwittler 2002.

für verschiedene Befragtengruppen mit unterschiedlichem kognitivem Niveau separat zu berechnen. Ein Vergleich der Effektgrößen für diese verschiedenen Teilpopulationen ermöglicht uns dann den Test der Hypothesen.

Zur Bildung der verglichenen Teilpopulationen wurde zunächst das Alter herangezogen. Unter Berücksichtigung der Fallzahlen in den einzelnen Jahrgängen haben wir die Befragten zu Altersgruppen zusammengefasst, die aus 2 bis 4 Jahren bestehen. Nur bei der altersmäßigen Kategorisierung der Befragten in Studie 3 haben wir – auf Grund der geringen Fallzahlen – größere Altersgruppen bilden müssen. Wegen der unterschiedlichen Altersspannen lassen sich die Altersgruppen der drei Studien nicht direkt vergleichbar gestalten – die Fallzahlen der einzelnen Alterskategorien wären sonst sehr ungleich und in einzelnen Fällen zu gering, um die Effekte sichtbar zu machen.

Das Niveau der kognitiven Kapazität haben wir weiter durch die schulischen Leistungen/Bildungsniveau gemessen (vgl. zu dieser Operationalisierung Knäuper 1999): In den Studien 1 und 2 basiert die Kategorisierung auf einer zusammenfassenden subjektiven Einschätzung der Zufriedenheit der Lehrer mit den schulischen Leistungen durch die Befragten;³ in Studie 3 wurde auf den allgemein bildenden Schulabschluss (den erreichten bzw. den Angestrebten) zurückgegriffen. Diese relativ indirekte Messung der kognitiven Kapazität hängt zum einen mit den hohen Ressourcenansprüchen einer direkten, umfassenden Messung zusammen (Platz im Fragebogen und Zeit im Interview), und zum anderen mit der Tatsache, dass das Konzept der kognitiven Kapazität in Bezug auf den Frage-Antwort-Prozess bisher nur ansatzweise theoretisch ausgearbeitet ist. Die Selbsteinstufung der schulischen Leistungen bzw. das allgemeine Bildungsniveau wurden zu drei Kategorien rekodiert („gut“ – bei den schulischen Leistungen – bzw. „hoch“ – beim Bildungsniveau –, „mittel“, „niedrig“)⁴ und mit Blick auf die Fallzahlen in den einzelnen Zellen für die Zwecke der hier vorgenommenen Analysen zu „gut“ bzw. „hoch“ einerseits und „mittel/niedrig“ andererseits zusammengefasst.

3 Dabei wurde „sehr zufrieden“ und „zufrieden“ als Indikator für „gute“ schulische Leistungen aufgefasst. Die übrigen Ausprägungen („teils/teils“, „weniger zufrieden“ und „gar nicht zufrieden“) wurden zu „mittleren/niedrigen“ schulischen Leistungen zusammengefasst.

4 Beim Bildungsniveau wurde die (fachgebundene und die allgemeine) Hochschulreife und der (Fach-)Hochschulabschluss als hohes Bildungsniveau codiert; alle übrigen Kategorien wurden zu „mittel/niedrig“ zusammengefasst (kein Abschluss, Hauptschulabschluss, Mittlere Reife).

3. Ergebnisse

Im Folgenden konzentrieren wir uns zunächst auf Antwortreihenfolgen-Effekte (3.1) bevor wir Skalen-Effekte (3.2) und die Effekte der numerischen Werte, die den Antwortkategorien zugeordnet sind, berichten (3.3). Im abschließenden Teilkapitel werden wir uns den Ergebnissen von zwei Experimenten zur Fragereihenfolge zuwenden (3.4.).

3.1 Antwortreihenfolge-Effekte

Experiment 1

Als Teil von Studie 2 wurden die Befragten gebeten, aus einer Liste von vorgegebenen Jugendkulturen diejenige auszuwählen, der sie sich am nächsten fühlen. In einem der beiden randomisierten Fragebogensplitts waren die zehn Antwortkategorien in vollständig gedrehter Reihenfolge abgedruckt (vgl. Abbildung 1).

Abbildung 1: Frageformulierung in Experiment 1 zur Antwortreihenfolge, Originalversion (links) und gedrehte Version (rechts)

26. Stell dir vor, es kommt ein neuer Mitschüler in deine Klasse und sieht dich das erste Mal. Was würde der denken welcher Gruppierung du nahe stehst? (bitte nur eine Antwort ankreuzen)

- Rapper (Hip-Hop)
- Techno
- Skinheads
- Autonome
- Alternative
- Gothics
- Metaller
- Rockabilly, Rock'n'Roll
- Hooligans
- Punks

zu einer anderen Gruppe: _____

26. Stell dir vor, es kommt ein neuer Mitschüler in deine Klasse und sieht dich das erste Mal. Was würde der denken welcher Gruppierung du nahe stehst? (bitte nur eine Antwort ankreuzen)

- Punks
- Hooligans
- Rockabilly, Rock'n'Roll
- Metaller
- Gothics
- Alternative
- Autonome
- Skinheads
- Techno
- Rapper (Hip-Hop)

zu einer anderen Gruppe: _____

Von Krosinck und Alwin (1987) ausgehend erwarteten wir einen Primacy-Effekt, denn in selbst-administrierten Befragungen erhalten Antwortmöglichkeiten, die am Anfang einer langen Liste von Kategorien präsentiert werden, mehr Aufmerksamkeit und sie setzen einen interpretativen Rahmen für die Wahrnehmung der nachfolgenden Antworten (im Sinn eines Standards für den Vergleich). Hinzu kommt, dass „a satisfying criterium implies that the earlier in the list an acceptable answer appears the more popular it will be. If respondents chose an earlier option they avoid thinking about the others” (Tourangeau et al. 2000: 250).

Tabelle 1: Größe des Antwortreihenfolgen-Effekts (Recency-Effekt, Prozentpunkte) nach Alter und schulischen Leistungen

Altersgruppe	Schulische Leistungen		Alle
	Gut	Mittel/Schlecht	
13 bis 15 Jahre	19 *	10 *	11 ***
16 bis 17 Jahre	10	2	4
18 bis 21 Jahre	-1	0	1
Alle	7	5	6 ***

Studie 2, Frage 26. Angegeben ist die Prozentpunktdifferenz = Anteil der Nennungen bei Positionierung an letzter Stelle minus Anteil der Nennungen bei Positionierung am Anfang; ohne die „anderen Gruppen“, die insgesamt nur selten genannt wurden (7,7%).

* $p < 0,05$; *** $p < 0,001$ für die jeweiligen Effekte; Test auf Differenzen zwischen den Effekten: Alter $p < 0,07$, schulische Leistungen nicht signifikant.

Da es sich um die Differenz von Prozentsatzdifferenzen handelt, sind Standard-Tests für die Unterschiede zwischen den Antworteffekten nicht verfügbar. Hier wird ein χ^2 -Anunabhängigkeitstest auf die Größe der Prozentsatzdifferenzen unter Berücksichtigung der Fallzahl in der entsprechenden Kategorie der Randverteilung berechnet.

Entgegen dieser Hypothese und den in der Literatur dokumentierten Ergebnissen (z.B. Hippler/Schwarz 1992) beobachten wir jedoch keinen Primacy-Effekt, sondern einen Recency-Effekt in unseren Daten: Die Antwortkategorien, die am Ende der Liste stehen, werden unabhängig vom Inhalt häufiger genannt (6 Prozentpunkte; $p < 0,001$).⁵ Nachdem es in unserer Studie jedoch vordringlich darum geht, die Größe des Antwortreihenfolgen-Effektes in Abhängigkeit vom Alter und von den schulischen Leistungen zu untersuchen, vernachlässigen wir möglichen Ursachen für das Auftreten dieses Recency-Effekts und analysieren stattdessen die Effektgröße in Abhängigkeit von diesen Variablen.

Ausgehend von den theoretischen Überlegungen, die wir im ersten Kapitel zusammengefasst haben, erwarten wir, dass ältere Befragte und Befragte mit besserer schulischer Leistung eher in der Lage sind alle Antwortkategorien zur gleichen Zeit im Gedächtnis zu behalten und daher zu einem größeren Anteil jeder Antwortkategorie die gleiche Aufmerksamkeit zuzuwenden. Dies sollte sich in einem weniger ausgeprägten Antwortreihenfolgen-Effekt ausdrücken. Wie erwartet variiert die Effektgröße mit dem Alter: Jüngere

5 Wir vermuten, dass der Recency-Effekt vor allem als Folge der Tatsache auftritt, dass die angebotenen Jugendkulturen für einen beträchtlichen Teil der Befragten wenig plausibel erscheinen. Sudman und Kollegen haben betont (1996, S. 141), dass „a given implausible item should elicit more disagreeing thoughts and be less likely to be endorsed when presented early rather than late in the list“.

Befragte weisen tendenziell größere Effekte als ältere Jugendliche und volljährige Befragte auf ($p < 0,07$). Die schulische Leistungsfähigkeit hat jedoch keinen signifikanten Einfluss auf die Größe der Antwort-Reihenfolgen-Effekte.

Experiment 2

Im zweiten Experiment zu Antwortreihenfolgen-Effekten (Studie 1) wurden die Befragten ebenfalls mit einer langen Liste von Antwortkategorien konfrontiert, in der infrastrukturelle Einrichtungen des Wohnquartiers angegeben waren (Abbildung 2).

Abbildung 2: Fragebogenversionen in Experiment 2 zur Antwortreihenfolgen, Originalversion (rechts) und gedrehte Version (links; die übrigen Versionen sind nicht abgedruckt)

- | | |
|--|--|
| <p>31. Welche der folgenden Einrichtungen gibt es in deinem Wohnviertel oder deinem Dorf? (Du kannst mehrere Antworten ankreuzen.)</p> <ul style="list-style-type: none"> <input type="checkbox"/> Supermarkt <input type="checkbox"/> Kinderspielplatz <input type="checkbox"/> Kindergarten <input type="checkbox"/> Museum <input type="checkbox"/> Theater <input type="checkbox"/> Bücherei <input type="checkbox"/> Busverbindungen <input type="checkbox"/> Schule <input type="checkbox"/> Hallenbad <input type="checkbox"/> Jugendzentrum <input type="checkbox"/> Diskothek <input type="checkbox"/> Kino | <p>31. Welche der folgenden Einrichtungen gibt es in deinem Wohnviertel oder deinem Dorf? (Du kannst mehrere Antworten ankreuzen.)</p> <ul style="list-style-type: none"> <input type="checkbox"/> Kino <input type="checkbox"/> Diskothek <input type="checkbox"/> Jugendzentrum <input type="checkbox"/> Hallenbad <input type="checkbox"/> Schule <input type="checkbox"/> Busverbindungen <input type="checkbox"/> Bücherei <input type="checkbox"/> Theater <input type="checkbox"/> Museum <input type="checkbox"/> Kindergarten <input type="checkbox"/> Kinderspielplatz <input type="checkbox"/> Supermarkt |
|--|--|

Die Befragten sollten aus dieser Liste diejenigen Ausstattungselemente auswählen, die in ihrer Wohnumgebung vorhanden sind. Die 12 Antwortkategorien wurden in vier verschiedenen Reihenfolgen präsentiert: Neben der Original-Version gab es eine vollständig gedrehte Version und zwei weitere Versionen, bei denen die Antwortkategorien vom Anfang und vom Ende der Liste an unterschiedlichen Positionen in der Mitte erschienen.

Wie erwartet (Krosnick/Alwin 1987; Schwarz et al. 1992) zeigen die Daten einen moderaten Primacy-Effekt. Antwortkategorien, die am Beginn der Liste erscheinen, erhalten unabhängig vom Inhalt der Antwortkategorie höhere Zustimmungswerte verglichen mit einer Positionierung in der Mitte oder am Ende der Liste. In Tabelle 2 werden die Ergebnisse für die Antwortkategorie „Kino“ (= es gibt ein Kino im Wohnquartier der Befragten) dargestellt. Wird „Kino“ am Anfang der Liste der Antwortkategorien präsentiert, entscheiden sich signifikant mehr Befragte dafür (5 Prozentpunkte, $p < 0,001$).

Tabelle 2: Größe des Antwortreihenfolgen-Effektes (Primacy-Effekt, Prozentpunkte) nach Alter und schulischen Leistungen

Altersgruppe	Schulische Leistungen		Alle
	Gut	Mittel/Schlecht	
10 bis 13 Jahre	7 *	11 *	10 ***
14 bis 17 Jahre	1	5	3
18 bis 21 Jahre	5	2	4
Alle	4 *	6 **	5 ***

Studie 1, Frage 31, Kategorie "Kino". Angegeben ist die Prozentsatzdifferenz zwischen der Originalversion und den drei übrigen Version; also Anteil der Nennungen bei Positionierung am Anfang minus Anteil der Nennungen bei Positionierung in der Mitte oder am Ende.

* < 0,05; ** p < 0,01, *** p < 0,001 für die jeweiligen Effekte; Test auf Unterschiede zwischen den Effekten: Alter p < 0,01, schulische Leistungen p < 0,05; vgl. die Anmerkung zu Tabelle 1.

Die Größe dieses Antwortreihenfolgen-Effektes korreliert wie erwartet mit dem Alter der Befragten: Jüngere Befragte zwischen 10 und 13 Jahren zeigen größere Effekte (10 Prozentpunkte) im Vergleich zu älteren Befragten ($p < 0,01$). Außerdem ist auch die schulische Leistungsfähigkeit korreliert: Für Schüler mit besseren schulischen Leistungen errechnet sich ein kleinerer Antwortreihenfolgen-Effekt (4 Prozentpunkte) als bei den Befragten mit durchschnittlichen oder schlechten schulischen Leistungen (6 Prozentpunkte; $p < 0,05$).

Insgesamt unterstützen die Befunde aus den Experimenten 1 und 2 die Behauptung, dass sich jüngere Befragte stärker von der Antwortreihenfolge leiten lassen, also aufgrund der geringeren kognitiven Verarbeitungskapazitäten den einzelnen Antwortmöglichkeiten unterschiedlich viel Aufmerksamkeit zuwenden. Der entsprechende Einfluss schlechter schulischer Leistungen lässt sich nur in Experiment 2 nachweisen.

3.2 Skalen-Effekte

Das von uns benutzte Experiment zu den Skalen-Effekten basiert auf einer klassischen Studie von Schwarz und Mitarbeitern (1985), in dem die Befragten gebeten wurden, die Anzahl der Stunden zu schätzen, die sie pro Tag durchschnittlich fernsehen. Die Befragten wurden dabei zufällig zwei verschiedenen Antwortschemata zugewiesen, die entweder von „bis zu einer ½ Stunde“ bis „mehr als 2½ Stunden“ reichten oder aber von „bis zu 2 ½ Stunden“ bis „mehr als 4 ½ Stunden“. In der von Schwarz und Mitarbeitern (1985) durchgeführten Studie berichteten Befragte, die mit der höherreichenden Antwortskala konfrontiert wurden, einen signifikant längeren durchschnittlichen TV-Konsum im Ver-

gleich zu den mit der niedrigeren Werte umfassenden Antwortskala konfrontierten Befragten.

Experiment 3

In dem in Studie 2 integrierten Experiment 3 haben wir diese klassische Studie mit Kindern und Jugendlichen repliziert. Die befragten Kinder und Jugendlichen waren gebeten, ihren täglichen TV-Konsum auf einer 6-stufigen Skala zu berichten, die analog zu den Kategorien von Schwarz und Mitarbeitern (1985) in zwei verschiedenen Varianten dargeboten waren. Entsprechend können wir voraussagen, dass Befragte, die mit den höheren Antwortkategorien konfrontiert werden, eine längere tägliche Fernsehdauer berichten werden, als Kinder und Jugendliche, die die Frage mit den niedrigen Antwortkategorien beantworten (= Skalen-Effekt).

Abbildung 3: Fragebogenversionen im Experiment 3 zum Skalen-Effekt, „täglicher TV-Konsum“

24. Wie viele Stunden am Tag schaust du ungefähr fern?

- bis 2¼ Stunde
- bis 3 Stunde
- bis 3½ Stunden
- bis 4 Stunden
- bis 4¼ Stunden
- über 4½ Stunden

24. Wie viele Stunden am Tag schaust du ungefähr fern?

- bis ½ Stunde
- bis 1 Stunde
- bis 1½ Stunden
- bis 2 Stunden
- bis 2½ Stunden
- über 2½ Stunden

Insgesamt beobachten wir – im Einklang mit den Befunden von Schwarz und Mitarbeitern (1985) – einen beträchtlichen Skalen-Effekt: Der Anteil der Befragten, die eine Antwortkategorie ausgewählt haben, die eine längere als 2½-stündige tägliche Fernsehdauer repräsentiert, ist um 25 Prozentpunkte höher ($p < 0,001$), wenn die Frage die hohen Antwortalternativen enthält.

Ausgehend von den theoretischen Überlegungen im Einleitungskapitel erwarten wir weiter, dass jüngere Befragte und Befragte mit schlechteren Schulleistungen sich stärker an den Antwortkategorien orientieren als ältere Jugendliche und Befragte mit besseren Schulleistungen: Weil ihre kognitiven Fähigkeiten sich erst noch entwickeln, sind erstere weniger gut in der Lage, die gewünschte Schätzung des täglichen TV-Konsums ohne Rückgriff auf das Antwortkontinuum vorzunehmen, als Befragte mit besser entwickelten kognitiven Fähigkeiten – entsprechend werden wir bei ihnen einen größeren Skalen-Effekt beobachten.

Tabelle 3: Größe des Skalen-Effekts (Prozentpunkte) nach Alter und schulischen Leistungen

Altersgruppe	Schulische Leistungen		Alle
	Gut	Mittel/Schlecht	
13 bis 15 Jahre	25 ***	32 ***	30 ***
16 bis 17 Jahre	19 ***	28 ***	25 ***
18 bis 21 Jahre	22 ***	17 ***	19 ***
Alle	22 ***	27 ***	25 ***

Studie 2, Frage 24. Angegeben ist die Prozentsatzdifferenz für den Anteil der Nennungen über 2½ Stunden Fernsehdauer in der Version mit den hohen Antwortkategorien minus des Anteils der Nennungen bei der Version mit den niedrigen Antwortkategorien.

*** $p < 0,001$ für den jeweiligen Effekt; Test auf Unterschiede der Effekte: Alter $p < 0,01$, schulische Leistungen $p < 0,05$; vgl. die Anmerkung zu Tabelle 1.

Entsprechend unserer Hypothese zeigen jüngere Befragte größere Skalen-Effekte als Ältere. Während der Skalen-Effekt bei den volljährigen Personen 19 Prozentpunkte erreicht, kommen Befragte zwischen 16 und 17 Jahren auf 25 Prozentpunkte und in der Weiterführung dieses Trends berechnen wir in der jüngsten Altersgruppe der bis 15-Jährigen einen Skalen-Effekt von 30 Prozentpunkten ($p < 0,01$). Außerdem zeigt sich, dass Personen mit schlechteren schulischen Leistungen größere Effekte aufweisen (27 Prozentpunkte) als solche mit besseren Leistungen (22 Prozentpunkte; $p < 0,05$). Diese Ergebnisse unterstützen die Hypothese wonach Befragte mit besseren kognitiven Fähigkeiten eher in der Lage sind, Antworten unabhängig vom angebotenen Antwortschema zu generieren. Befragte mit weniger entwickelten kognitiven Fähigkeit orientieren sich entsprechend stärker am Hinweischarakter der Antwortkategorien. Eine weitere Inspektion der Ergebnisse zeigt weiter, dass es vor allem die jüngeren, weniger erfolgreichen Schüler sind, die den größten Skalen-Effekt aufweisen (32 Prozentpunkte).

Experiment 4

In einem zweiten Experiment zu Skalen-Effekten haben wir das gleiche klassische Design von Schwarz und Mitarbeitern (1985) benutzt, wie im vorangegangenen Experiment. Die Befragten in Studie 3 wurden wiederum gebeten, ihren täglichen Fernsehkonsum auf einer 6-stufigen Skala zu schätzen. Dabei haben wir die beiden experimentellen Versionen des Antwortkontinuums verwendet, das bereits oben ausführlich beschrieben wurde (vgl. Experiment 3 und Abbildung 3 für Details des Designs). Wie bereits zuvor beschrieben, erwarten wir, dass die mit den höheren Antwortkategorien konfrontierten Befragten eine

längere tägliche Fernsehdauer berichten, als diejenigen, die die Frage anhand der niedrigeren Antwortkategorien beantwortet. Unsere Ergebnisse bestätigen im Wesentlichen diese Voraussage: Insgesamt ist der Anteil der Befragten, die einen täglichen TV-Konsum von mehr als 2½ Stunden berichten, um 31 Prozentpunkte höher, wenn sie die höheren Antwortkategorien benutzen ($p < 0,001$).

Tabelle 4: Größe des Skalen-Effekts (Prozentpunkte), nach Alter und Bildungsniveau

Altersgruppe	Schulische Leistungen		Alle
	Gut	Mittel/Schlecht	
15 bis 20 Jahre	24 +	37 ***	34 ***
21 bis 60 Jahre	20 **	25 **	26 ***
61 bis 90 Jahre	21 +	41 ***	35 ***
Alle	22 ***	33 ***	31 ***

Studie 3, Frage 6-2. Angegeben ist die Prozentsatzdifferenz für den Anteil der Nennungen über 2½ Stunden Fernsehdauer in der Version mit den hohen Antwortkategorien minus des Anteils der Nennungen bei der Version mit den niedrigen Antwortkategorien.

+ < 0,10; ** $p < 0,01$; *** $p < 0,001$ für die jeweiligen Effekte; Test auf Unterschiede zwischen den Effekten: Alter nicht signifikant, Bildungsniveau $p < 0,001$; vgl. die Anmerkung zu Tabelle 1.

Studie 3 ermöglicht uns eine vergleichende Analyse von jüngeren Befragten zwischen 15 und 20 Jahren und älteren Befragten zwischen 61 und 90 Jahren und zudem einen Vergleich nach Bildungsniveau. Testet man diesen Effekt unter Berücksichtigung der besuchten bzw. absolvierten Schulart, die hier anstelle der schulischen Leistungen als Indikator verwendet wurde, um auch die Erwachsenen einbeziehen zu können, so ergibt sich ein deutlicher Effekt ($p < 0,001$): Befragte in niedrigen Bildungsgängen (oder Erwachsene, die niedrige Bildungsgänge absolviert haben) weisen einen signifikant größeren Skalen-Effekt auf (33 Prozentpunkte) als Befragte, die sich in höheren Bildungsgängen befinden bzw. diese absolviert haben (22 Prozentpunkte). Außerdem scheint der Skalen-Effekt für die jungen und die alten Befragten größer ausgeprägt zu sein, als in der Erwachsenenpopulation zwischen 21 und 60 Jahren (nicht signifikant). Scheinbar verfügen junge und alte Befragte über geringere kognitive Ressourcen, und weisen daher größere Skalen-Effekte auf; allerdings sind diese Unterschiede nicht signifikant.

3.3 Numerische Werte, die mit den Antwortkategorien assoziiert sind

Experiment 5

In einem weiteren Experiment (Studie 2) haben wir den Einfluss der numerischen Werte, die den einzelnen Antwortkategorien einer Einstellungsfrage zugeordnet sind, auf das Antwortverhalten überprüft. Wir nehmen an, dass die Befragten die den Antwortalternativen beigefügten numerischen Werte verwenden, um ein besseres Verständnis der zu Grunde liegenden evaluativen Dimension der Frage zu entwickeln. In diesem Experiment, das auf Schwarz und Mitarbeiter (1991a) zurückgeht, wurden die Befragten gebeten, den „Erfolg in ihrem Leben“ auf einer 11-stufigen Skala (vgl. Abbildung 4) zu schätzen. Die Endpunkte der Skala waren mit „überhaupt nicht erfolgreich“ und „sehr erfolgreich“ verbalisiert. Die übrigen Skalenpunkte waren nicht verbal ausgezeichnet, sondern nur durch numerische Werte markiert. Zwei verschiedene Varianten dieser Frage wurden zufällig Teilgruppen des Gesamtsamples präsentiert. Eine Hälfte der Befragten erhielt einen Fragebogen mit numerischen Werten von -5 bis +5 und die andere Hälfte der Befragten beantwortete die gleiche Frage unter Benutzung der Werte von 0 bis 10.

Basierend auf den in der Literatur berichteten Ergebnissen (Schwarz et al. 1991a) erwarteten wir eine signifikante Verschiebung der Bewertung des Erfolgs im Leben in Abhängigkeit von den verwendeten numerischen Werten. In einem selbst-administrierten Fragebogen werden die Befragten das Antwortkontinuum von -5 bis +5 derart interpretieren, dass der Wert 0 den neutralen Mittelpunkt (nicht erfolgreich) repräsentiert, wohingegen -5 das Gegenteil von Erfolg, also z.B. Misserfolg symbolisiert (Schwarz/Hippler 1995). Im Vergleich werden Befragte, die die gleiche Frage mit den Werten 0 bis 10 präsentiert bekommen, den linken extremen Punkt (also den Wert 0) als Abwesenheit von Erfolg und nicht als Misserfolg interpretieren. Daher sollte die Einschätzung des Erfolgs im Leben bei Befragten, die mit einer Skala von -5 bis +5 konfrontiert sind, in Richtung auf das positive Ende der Skala verschoben sein.

Abbildung 4: Fragebogenversionen in Experiment 5 zum Effekt der mit den Antwortalternativen assoziierten numerischen Werte, „Erfolg im Leben“

13. Was würdest du sagen: wie erfolgreich warst du in deinem Leben bisher alles in allem?

<input type="checkbox"/> -5	<input type="checkbox"/> -4	<input type="checkbox"/> -3	<input type="checkbox"/> -2	<input type="checkbox"/> -1	<input type="checkbox"/> 0	<input type="checkbox"/> +1	<input type="checkbox"/> +2	<input type="checkbox"/> +3	<input type="checkbox"/> +4	<input type="checkbox"/> +5
überhaupt nicht erfolgreich					0					sehr erfolgreich

13. Was würdest du sagen: wie erfolgreich warst du in deinem Leben bisher alles in allem?

<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	<input type="checkbox"/> 7	<input type="checkbox"/> 8	<input type="checkbox"/> 9	<input type="checkbox"/> 10
überhaupt nicht erfolgreich										sehr erfolgreich

Tabelle 5 fasst die Ergebnisse dieses Experiments zusammen. Insgesamt beobachten wir eine nennenswerte Verschiebung der Antwortverteilung in Abhängigkeit von den verwendeten numerischen Werten. Wenn Befragte mit der Skala von -5 bis +5 konfrontiert werden, ist der Anteil der Befragten, die eine Antwort oberhalb des Mittelpunktes wählen, um 13 Prozentpunkte höher ($p < 0,001$) – analog zu den in der Literatur berichteten Befunden (Schwarz et al. 1991a; Schwarz/Hippler 1995).

Auch hier benutzen wir die Größe dieses Effekts als abhängige Variable, um den Einfluss von Alter und schulischen Leistungen auf den Frage-Antwort-Prozess zu untersuchen. Weil wir vermuten, dass jüngere Kinder und Schüler mit schlechteren schulischen Leistungen geringer ausgeprägte kognitive Fähigkeiten aufweisen, rechnen wir bei ihnen mit einem größeren Effekt. Denn diese Befragten sind weniger in der Lage, der komplizierten Aufgabe beim Aufbau einer kognitiven Repräsentation der evaluativen Dimension dieser Frage allein anhand der verbalisierten Endpunkte des Antwortkontinuums nachzukommen. Daher werden sie sich bei der Entwicklung des Frageverständnisses und der Generierung einer Antwort stärker von den numerischen Werten beeinflussen lassen.

Tabelle 5: Effekte der den Antwortalternativen beigefügten numerischen Werte (Prozentpunkte), nach Alter und schulischen Leistungen

Altersgruppe	Schulische Leistungen		Alle
	Gut	Mittel/Schlecht	
13 bis 15 Jahre	9 *	18 ***	16 ***
16 bis 17 Jahre	8 *	13 ***	12 ***
18 bis 21 Jahre	8 **	12 ***	11 ***
Alle	9 ***	15 ***	13 ***

Studie 2, Frage 13. Angegeben ist die Prozentsatzdifferenz: Anteil der Befragten über dem Mittelpunkt der Skala in der Version -5/+5 minus des gleichen Anteils in der Version 0/10.

* < 0,05; *** $p < 0,001$ für die jeweiligen Effekte; Test auf Unterschiede zwischen den Effekten: Alter $p < 0,01$, schulische Leistungen $p < 0,001$; vgl. die Anmerkung zu Tabelle 1.

Wie erwartet weisen ältere Befragte einen geringeren Antworteffekt auf (11 Prozentpunkte) als jüngere Befragte (16 Prozentpunkte; $p < 0,01$). Weiterhin beobachten wir einen Unterschied zwischen Befragten mit mittleren und schlechten schulischen Leistungen (15 Prozentpunkte) und Schülern mit besserer schulischer Leistungsfähigkeit (9 Prozentpunkte; $p < 0,001$). Beide Befunde gehen mit den theoretischen Erwartungen konform. Eine detaillierte Inspektion der Tabelle ergibt weiter, dass wir vor allem für die Befragten mit durchschnittlicher und schlechterer schulischer Leistungsfähigkeit einen

ausgeprägten Alterseffekt nachweisen können. Hingegen finden wir keinen Alterseffekt in der Gruppe mit überdurchschnittlichen schulischen Leistungen.

3.4 Fragereihenfolgen-Effekte

Bisher haben geringere kognitive Fähigkeiten zu stärkeren Antworteffekten geführt (Antwortreihenfolgen-Effekte, Skalen-Effekte und Effekte der numerischen Werte an den Antwortalternativen). Im Vergleich dazu wirkt die kognitive Kapazität mit umgekehrtem Vorzeichen auf die Höhe der Fragereihenfolgen-Effekte. Da diese auf der Berücksichtigung von zuvor gestellten Fragen bei der Interpretation und Beantwortung der aktuellen Frage beruhen, müssen Befragte, die zuvor wahrgenommenen Informationen speichern und als relevant für die Beantwortung der aktuellen Frage wahrnehmen. Diese kognitive Aufgabe können jüngere Befragte und Befragte mit schlechteren schulischen Leistungen weniger gut bewältigen, weshalb der Inhalt der zuvor gestellten Frage bei ihnen eine geringere Rolle für das Frageverständnis der aktuellen Frage und deren Beantwortung spielt. Entsprechend sind die Fragereihenfolgen-Effekte bei den Befragten mit geringeren kognitiven Kapazitäten geringer ausgeprägt.

Experiment 6

Im ersten Experiment zur Fragereihenfolge (Studie 2) wurden die Befragten gebeten, ihre Chancen auf einen Arbeitsplatz nach der Schule im Vergleich zu anderen Jugendlichen in Ostdeutschland und Westdeutschland zu schätzen. Dabei wurden jedem Befragten (es handelt sich nur um Westdeutsche im Bundesland Bayern) zwei Fragen gestellt: In der ersten sollten die Chancen im Vergleich zu anderen Jugendlichen in Westdeutschland geschätzt werden; in der zweiten Frage sollte ein entsprechender Vergleich mit Jugendlichen in Ostdeutschland vorgenommen werden (Abbildung 5). Diese beiden Fragen wurden in der Originalreihenfolge (der Vergleich mit Jugendlichen in Westdeutschland zuerst) und in umgekehrter Reihenfolge (der Vergleich mit ostdeutschen Jugendlichen zuerst) gestellt.

Abbildung 5: Fragebogenversionen in Experiment 6 zur Fragereihenfolge, „Chancen auf einen Arbeitsplatz nach der Schule“, Originalversion (rechts) und gedrehte Version (links)

11. Und wenn du dich mit den Jugendlichen in OST-Deutschland allgemein vergleichst, hast du dann deiner Meinung nach bessere Chancen auf einen Arbeitsplatz, schlechtere Chancen oder etwa die gleichen Chancen auf einen Arbeitsplatz?

- bessere Chancen
- etwa die gleichen Chancen
- schlechtere Chancen

12. Und wenn du dich mit den Jugendlichen in WEST-Deutschland allgemein vergleichst, hast du dann deiner Meinung nach bessere Chancen auf einen Arbeitsplatz, schlechtere Chancen oder etwa die gleichen Chancen auf einen Arbeitsplatz?

- bessere Chancen
- etwa die gleichen Chancen
- schlechtere Chancen

11. Und wenn du dich mit den Jugendlichen in WEST-Deutschland allgemein vergleichst, hast du dann deiner Meinung nach bessere Chancen auf einen Arbeitsplatz, schlechtere Chancen oder etwa die gleichen Chancen auf einen Arbeitsplatz?

- bessere Chancen
- etwa die gleichen Chancen
- schlechtere Chancen

12. Und wenn du dich mit den Jugendlichen in OST-Deutschland allgemein vergleichst, hast du dann deiner Meinung nach bessere Chancen auf einen Arbeitsplatz, schlechtere Chancen oder etwa die gleichen Chancen auf einen Arbeitsplatz?

- bessere Chancen
- etwa die gleichen Chancen
- schlechtere Chancen

Wir erwarten, dass die Fragereihenfolge einen moderierenden Effekt auf die perzipierte Bedeutung der Fragen ausübt: Vergleichen Befragte sich selbst zunächst mit Jugendlichen in Ostdeutschland, sollten beim anschließenden Vergleich mit Jugendlichen in Westdeutschland bessere Einschätzungen gemessen werden. Wir vermuteten, dass die ggf. wahrgenommene Unterlegenheit der eigenen Person im Vergleich zu den westdeutschen Altersgenossen vor dem Hintergrund der größeren Distanz zu den Ostdeutschen kleiner erscheint; die eigene Lage wird dadurch positiver berichtet.

Tabelle 6: Größe des Fragenreihenfolgen-Effekts, „Chance auf einen Arbeitsplatz nach der Schule“ (Prozentpunkte), nach Alter und schulischen Leistungen

Altersgruppe	Schulische Leistungen		Alle
	Gut	Mittel/Schlecht	
13 bis 15 Jahre	2	0	0
16 bis 17 Jahre	7	-1	1
18 bis 21 Jahre	13	6	10
Alle	7 **	5 **	6 **

Studie 2, Frage 11. Angegeben ist die Prozentsatzdifferenz des Anteils der Befragten, die eine gute Chance auf einen Arbeitsplatz erwarten in der gedrehten Version minus des entsprechenden Anteils in der Originalversion.

* $p < 0,05$; ** $p < 0,01$ für den jeweiligen Effekt; Test auf Unterschiede zwischen den Effekten: Alter $p < 0,01$, schulische Leistungen $p < 0,11$; vgl. die Anmerkung zu Tabelle 1.

Insgesamt finden wir einen moderaten erwartungskonformen Fragereihenfolgen-Effekt von 6 Prozentpunkten ($p < 0,01$): Wenn die befragten Kinder und Jugendlichen sich zunächst mit den ostdeutschen Altersgenossen vergleichen und dann erst mit den Westdeutschen, dann ist der Anteil derjenigen, die sich selbst im Vergleich mit den Westdeutschen bessere Chancen auf einen Arbeitsplatz zumessen, um 6 Prozentpunkte höher, als wenn der Vergleich mit den Ostdeutschen nach dem Vergleich mit den Westdeutschen erfragt wird. Auch hier benutzen wir die Größe dieses Effektes als abhängige Variable und setzen sie mit der kognitiven Kapazität in Beziehung: Bei älteren Befragten zeigen sich signifikante Unterschiede der Beantwortung der Frage nach den Chancen auf einen Arbeitsplatz im Vergleich zu westdeutschen Jugendlichen, wohingegen bei jüngeren Befragten kein signifikanter Reihenfolgen-Effekt ausweisbar ist ($p < 0,01$). Der Effekt der schulischen Leistungen deutet zwar in die vorausgesagte Richtung, erreicht aber nicht das statistische Signifikanzniveau ($p < 0,11$).

Experiment 7

In einem zweiten Fragereihenfolgen-Experiment (Studie 2) wurden die Befragten gebeten, ihre Zufriedenheit mit ihrem Leben im Allgemeinen und ihre Zufriedenheit mit ihrer Situation in der Klasse anzugeben. In diesem Experiment haben wir die Logik eines klassischen Designs von Schwarz und Mitarbeiter (1991b; vgl. auch Sudman et al. 1996) auf jugendliche Befragte zugeschnitten. Je die Hälfte der Befragten erhielt eine von zwei experimentellen Fragebogenversionen zugewiesen: Version 1 enthielt die beiden Fragen in

der Originalreihenfolge, in der zunächst nach der Zufriedenheit mit dem Leben allgemein und dann erst nach der Zufriedenheit mit der Klasse gefragt wurde. Version 2 stellte die beiden Fragen in umgekehrter Reihenfolge (Abbildung 6). Wir erwarten, dass die Beurteilung der Situation in der Klasse negativer ausfällt, wenn die Frage an zweiter Stelle positioniert ist. Wenn die Kinder und Jugendlichen zunächst nach der Zufriedenheit mit dem Leben insgesamt gefragt werden und dann erst nach der Zufriedenheit mit ihrer Klasse, dann wird die Klasse im Kontrast zum Leben allgemein stärker als der Ort des Lernens und der Prüfungen aufgefasst, und weniger als Ort, wo man Freunde treffen oder gesellig sein kann. Wird hingegen nach der Klasse an erster Stelle gefragt, interpretieren die Befragten die Klasse als Zweiklang aus formalisierten Lernprozessen und sozialen Aktivitäten. Dies führt zu einer besseren Bewertung der Klasse, wenn die betreffende Frage an erster Stelle steht.

Abbildung 6: Fragebogenversionen in Experiment 7 zur Fragereihenfolge, „Zufriedenheit mit der Klasse“, Originalversion (links) und gedrehte Version (rechts)

1. Wie zufrieden bist du mit deinem Leben im Allgemeinen?

- Ich bin sehr zufrieden
- Ich bin eher zufrieden
- Ich bin eher unzufrieden
- Ich bin sehr unzufrieden

2. Wie zufrieden bist du mit deiner Klasse?

- Ich bin sehr zufrieden
- Ich bin eher zufrieden
- Ich bin eher unzufrieden
- Ich bin sehr unzufrieden

1. Wie zufrieden bist du mit deiner Klasse?

- Ich bin sehr zufrieden
- Ich bin eher zufrieden
- Ich bin eher unzufrieden
- Ich bin sehr unzufrieden

2. Wie zufrieden bist du mit deinem Leben im Allgemeinen?

- Ich bin sehr zufrieden
- Ich bin eher zufrieden
- Ich bin eher unzufrieden
- Ich bin sehr unzufrieden

Insgesamt finden wir in den Daten einen moderaten Fragereihenfolgen-Effekt von 7 Prozentpunkten ($p < 0,001$). Befragte, die zunächst nach der Klasse und dann erst nach ihrem Leben insgesamt befragt wurden, beurteilen die Situation in der Klasse entsprechend besser, als wenn sie die Fragen in umgekehrter Reihenfolge präsentiert bekommen.

Tabelle 7: Größe des Fragereihenfolgen-Effekts, „Zufriedenheit mit der Klasse“ (Prozentpunkte), nach Alter und schulischer Leistung

Altersgruppe	Schulische Leistungen		Alle
	Gut	Mittel/Schlecht	
13 bis 15 Jahre	9 *	4 *	5 *
16 bis 17 Jahre	6 +	8 ***	7 ***
18 bis 21 Jahre	11 ***	7 *	9 ***
Alle	8 ***	7 ***	7 ***

Studie 2, Frage 2 (ohne Berufsschüler). Angegeben ist die Prozentsatzdifferenz zwischen dem Anteil der Befragten, die sich in der gedrehten Version „sehr zufrieden“ oder „eher zufrieden“ mit der Klasse zeigen minus des entsprechenden Anteils in der Originalversion.

+ < 0,10; * $p < 0,05$; *** $p < 0,001$ für den jeweiligen Effekt; Test auf Unterschiede zwischen den Effekten: Alter $p < 0,001$, schulische Leistungen nicht signifikant; vgl. die Anmerkung zu Tabelle 1.

Auch diesen Fragereihenfolgen-Effekt haben wir nach Altersgruppen und schulischer Leistung differenziert und kommen zu folgenden Ergebnissen: Jüngere Befragte weisen im Vergleich zu älteren Jugendlichen einen signifikant kleineren Fragereihenfolgen-Effekt auf ($p < 0,001$). Außerdem können wir für Schüler mit besseren schulischen Leistungen leicht größere Fragenreihenfolgen-Effekte berechnen – allerdings sind diese Differenzen nicht statistisch signifikant.

Nimmt man die Ergebnisse der beiden Experimente zur Fragereihenfolge zusammen, so unterstützen die Befunde die Hypothese, wonach sich mit steigendem Alter der Umfang des relevanten Kontextes, der bei der Beantwortung einer Frage berücksichtigt wird, vergrößert. Nicht bestätigt werden kann hingegen die These, wonach die intellektuellen Fähigkeiten innerhalb einer Altersgruppe (gemessen durch die schulischen Leistungen) auf den Umfang des wahrgenommenen Kontextes einer Frage einwirken.

4. Diskussion

Betrachtet man die Ergebnisse aller Experimente zusammen, so zeigt sich ein durchgängiger Alterseffekt auf die Antworteffekte in der erwarteten Richtung, der mit Ausnahme von Experiment 4 auch statistisch gesichert ist. Sowohl bei der Antwortreihenfolge, bei den Skalen-Effekten und bei den Effekten der numerischen Werte sowie bei der Fragereihenfolge finden sich die erwarteten alterabhängigen Unterschiede in der Effektgröße. Mit Einschränkungen gilt dies auch für die schulischen Leistungen: Hier weisen die beobachteten Effekte – mit Ausnahme von Experiment 1 – in die durch die Hypothesen vorausgesagte Richtung. Außerdem ergibt die statistische Prüfung bei den

sagte Richtung. Außerdem ergibt die statistische Prüfung bei den Skalen-Effekten, bei den Effekten der numerischen Werte an den Antwortalternativen und bei der Antwortreihenfolge – bei einem von zwei Experimenten – gesicherte Unterschiede zwischen den Leistungsgruppen. Lediglich bei der Fragereihenfolge ergeben sich keine statistisch signifikanten Befunde.

Tabelle 8: Ergebnisse im Überblick, „Unterscheiden sich die Effekte in der erwarteten Richtung?“, Signifikanz der Effektunterschiede

	Indikatoren für die kognitive Kapazität	
	Alter	Schulische Leistungen
Effekt der Antwortreihenfolge		
Experiment 1	Ja +	Nein
Experiment 2	Ja **	Ja *
Skalen-Effekt		
Experiment 3	Ja **	Ja *
Experiment 4	Ja	Ja ***
Effekt der numerischen Werte		
Experiment 5	Ja **	Ja ***
Effekt der Fragereihenfolge		
Experiment 6	Ja **	Ja
Experiment 7	Ja **	Ja

+ p < 0,10; * p < 0,05; ** p < 0,01; *** p < 0,001.

Inwieweit die wenigen nicht erwartungskonformen Befunde auf methodische Schwächen unseres Designs oder auf falsche Hypothesen zurückzuführen sind, kann an dieser Stelle nicht entschieden werden. Die im Großen und Ganzen die Hypothese bestätigenden Ergebnisse ermutigen uns aber, in weiteren experimentellen Studien den Einfluss der kognitiven Fähigkeiten von Kindern und Jugendlichen auf die Beantwortung von Fragebögen zu untersuchen.⁶ Die vorliegenden Ergebnisse bilden zwar noch kein umfassendes Bild der kognitiven Prozesse dieser Befragtengruppen bei der Beantwortung von standardisierten Fragebögen, aber es ergibt sich eine Reihe von Hinweisen:

⁶ Derzeit ist eine weitere Untersuchung mit 4.500 befragten Kindern und Jugendlichen im Alter zwischen 10 und 21 Jahren im Feld, in die wir wiederum Experimente zu Antwort-Effekten integriert haben. Zudem haben wir versucht, die unabhängige Variable „schulische Leistungen“ durch Schulnoten in drei Hauptfächern zu erheben. Ergebnisse dieser Experimente liegen noch nicht vor.

Kinder und jüngere Jugendliche im Alter von 10 bis etwa 14 Jahre entwickeln aufgrund ihrer noch limitierten kognitiven Fähigkeiten ein eingeschränktes Frageverständnis; gleichzeitig lassen sie sich bei der Beantwortung der Frage stärker durch die vorgegebenen Antwortalternativen leiten. Es kompensieren sich hier also zwei gegenläufige Prozesse. Überspitzt könnte man sagen: Die jüngeren Befragten beantworten auch komplizierte Fragen, weil sie ihre noch limitierten kognitiven Fähigkeiten davor bewahrten zu erkennen, dass diese Fragen für sie ein Problem darstellen sollten. Dies mag im Übrigen auch eine Erklärung sein, warum der Item-Nonresponse bei Kindern zwar stärker ausgeprägt ist als bei Jugendlichen, aber auch kein übermäßig großes Problem darstellt (Borgers/Hox 2001; Fuchs 2003). Ältere Befragte betten Einzelfragen hingegen stärker in den Fragefluss bzw. -kontext ein, sie entwickeln ein breiteres Bild des Frageinhalts und sind daher bei der Generierung einer Antwort in einer anderen (besseren) Situation: Sie können die Antwort ohne (oder mit geringerem) Rückgriff auf die Spannweite des vorgegebenen Antwortkontinuums und auf die den Antwortvorgaben beigefügten numerischen Werte produzieren, und sie können allen Antwortalternativen die gleiche Aufmerksamkeit widmen. Damit scheint sich zumindest ansatzweise unsere generelle Annahme zu bestätigen, dass jüngere Befragte im Verhältnis zum Fragebogen in einer schwächeren, weniger autonomen Position verbleiben. Überspitzt könnte man sagen, dass jüngere Befragte einen möglichst stromlinienförmigen Weg durch die Befragung suchen, ohne eigene Spuren (= Anhaltspunkte für den „wahren Wert“) zu hinterlassen. Ob es sich dabei um ein Verhalten handelt, das Krosnick (1991, 1992) als „Satisficing“ bezeichnet hat, oder ob die befragten Kinder einfach nur versuchen, bei der anspruchsvollen Aufgabe der Beantwortung eines Fragebogens möglichst gut zu kooperieren (also versuchen, den Fragebogen auch bei Unklarheiten oder Problemen möglichst vollständig auszufüllen, vgl. Holaday/Turner-Henson 1989), kann an dieser Stelle noch nicht entschieden werden.

Einige wichtige Einschränkungen sind bei der Interpretation der Ergebnisse zu berücksichtigen: In unseren Analysen wurden Alter und schulische Leistungen als Indikatoren für die kognitive Kapazität verwendet. Wir sind uns jedoch bewusst, dass der Prozess der Entwicklung der kognitiven Kapazitäten nicht einer linearen altersbedingten Logik folgt. Vielmehr mögen manche Kinder bereits in jüngeren Jahren ein kognitives Stadium vorstoßen, dass andere erst zu einem späteren Zeitpunkt in ihrem Leben erreichen (Dockrell et al. 2000). Außerdem stehen Bildung und schulische Leistung nicht nur für kognitive Kapazität, sondern könnten ebenso Motivation und Leistungsbereitschaft repräsentieren. Hier kann erst eine Weiterentwicklung des Konzeptes der kognitiven Kapazität und seiner Operationalisierung zur Klärung beitragen (vgl. die Arbeiten zu den kognitiven Kapazitäten von alten und hochbetagten Befragten: Knäuper et al 1997; Knäuper 1999;

Schwarz/Knäuper 1999). Zu denken ist hierbei etwa an separate Messungen von Arbeitsgedächtnis und Aufmerksamkeit (vgl. Schwarz et al. 1998).

Weiter haben wir mit den referierten Ergebnissen nur einzelne Ausschnitte des Frage-Antwort-Prozesses untersuchen können. Insbesondere die Phase des Formatierens und des Editierens ist bisher weitgehend unberücksichtigt geblieben. Zudem stellt sich die Frage, ob die hier berichteten Unterschiede zwischen Kindern, Jugendlichen und volljährigen Befragten auch bei anderen Themen und mit anderen Befragungsmodes (mündlich-persönliche oder telefonische Befragung) repliziert werden können. Eine Generalisierung auf Befragungen von Kindern aller Art ist daher nicht zulässig.

Aus der jeweiligen Stichprobe der zugrunde liegenden drei Studien ergab sich für unsere Experimente eine Altersbegrenzung auf Kinder ab 10 Jahre. Eine Übertragung unserer Befunde auf Kinder jüngeren Alters ist daher nicht gerechtfertigt. Das Interesse der Umfrageforschung an Grundschulkindern wird aber vermutlich steigen (vgl. etwa das DJI-Kinderpanel), daher wäre eine Ausdehnung unserer Forschungslinie auf jüngere Kinder sinnvoll.

Weiter steht außer Frage, dass sich mit zunehmendem Alter auch die sozialen Kompetenzen der Kinder – und damit ihr Verständnis für die spezifische Situation „Befragung“ – vergrößert. Kohlberg (1976) hat die Bedeutung der sozialen Kompetenzen betont und in zukünftigen Studien sollten die kognitiven und sozialen Kompetenzen separat voneinander gemessen und als unabhängige Variablen benutzt werden.

In einem weiteren Schritt wäre es sinnvoll, die kognitiven Prozesse von Kindern und Jugendlichen mit denen von Erwachsenen an einem einheitlichen Datensatz zu vergleichen. Die Ergebnisse von Experiment 4 legen die Vermutung nahe, dass jüngere Befragte allgemein stärker von Antworteffekten betroffen sind als ein Querschnitt der Erwachsenenpopulation. Für eine detaillierte Analyse dieses Unterschieds fehlen uns jedoch bisher die Daten und wir sollten daher zukünftige Experimente so anlegen, dass Kinder, Jugendliche und Erwachsene miteinander vergleichbar sind und wir anhand der Größe der Antworteffekte entscheiden können, ab welchem Alter sich Kinder und Jugendliche bei der Beantwortung eines Fragebogens wie Erwachsene verhalten.

Weiter sollten Merkmale der einzelnen Fragen berücksichtigt werden, um aufzuzeigen, bei welchen Fragetypen die begrenzten kognitiven Kapazitäten der Kinder und jüngeren Jugendlichen zu Schwierigkeiten führen: Unsere Befunde zum Einfluss von Eigenschaften der jeweiligen Frage auf den Item-Nonresponse bei Kindern und Jugendlichen unterschiedlichen Alters und unterschiedlichen schulischen Leistungsniveaus (Fuchs 2003) verdeutlichen, dass die Probleme des Item-Nonresponse vor allem dann auftreten, wenn

die Fragen aufgrund der formalen Eigenschaften (z.B. Mehrfachantwort, Item-Batterie) oder aufgrund von inhaltlichen Anforderungen als schwierig einzustufen sind. In zukünftigen Experimenten sollten daher die hier geschilderten Experimente für Fragen auf unterschiedlichem Schwierigkeitsniveau repliziert werden. Auch Verständnisschwierigkeiten mit den Frageformulierungen sollten untersucht werden, denn möglicher Weise sind die beobachteten Effekte auch auf nicht altersgemäße Frageformulierungen zurückzuführen.

Trotz dieser Einschränkungen und der weiteren Forschungsfragen geben unsere Befunde Hinweise darauf, dass sich Kinder und Jugendliche hinsichtlich der Abarbeitung des Frage-Antwort-Prozesses von erwachsenen Befragten unterscheiden. Um daraus Konsequenzen für die Fragebogenkonstruktion zu ziehen, ist es derzeit wohl noch zu früh, weil wir nur über ein begrenztes Verständnis der Besonderheiten des Frage-Antwort-Prozesses bei Kinder und Jugendlichen verfügen. Aber zumindest scheint gesichert, dass wir den für erwachsene Befragte erarbeiteten methodischen Kenntnisstand nicht ohne weitere Prüfung auf Kinder und Jugendliche übertragen können. Das klingt wie eine methodische Binsenweisheit, aber zumindest erhalten wir auf Basis der referierten Ergebnisse eine Ahnung von Art und Umfang der Unterschiede bei der Verarbeitung von Fragebogenfragen.

Literatur

Amato, P./Ochiltree, G., 1987: Interviewing children about their families: a note on data quality. In: *Journal of Marriage and the Family*, 49, 669-675.

Benson, J./Hocevar, D., 1985: The impact of item phrasing on the validity of attitude scales for elementary school children. In: *Journal of Educational Measurement*, 22, 231-240.

Borgers, N./Hox, J., 2000: Reliability of responses in self-administered questionnaires: Research with children. In: Blasius, J./Hox, J./de Leeuw, E./Schmidt, P. (Hrsg.), *Social Science Methodology in the New Millennium. Proceedings of the Fifth International Conference on Logic and Methodology*, October 3-6, Cologne, Germany.

Borgers, N./Hox, J., 2001: Item nonresponse in questionnaire research with children. In: *Journal of official statistics*, 17, 321-335.

Borgers, N./de Leeuw, E./Hox, J., 2000: Children as respondents in survey research: Cognitive development and response quality. In: *Bulletin de Methodologie Sociologique (BMS)*, 66, 2000, 60-75.

Dockrell, J./Lewis, A./Lindsay, G., 2000: Researching childrens' perspectives: a psychological dimension. In: Lewis, A./Lindsay, G. (Hrsg.), *Researching childrens' perspectives*. Buckingham, 46-58.

Fuchs, M., 2003: Data quality in surveys among the youth. Item non-response and response errors. Vortrag anlässlich der internationalen Fachtagung "Methoden und Techniken der Kindheitsforschung", Schloss Hohenkammer, April 2003.

Fuchs, M./Lamnek, S./Luedtke, J., 1999: *Tatort Schule. Gewalt an Schulen 1994 – 1999*. Opladen.

Fuchs, M./Lamnek, S./Wiederer, R., 2001: *Querschläger. Jugendliche zwischen rechter Ideologie und Gewalt*. Opladen.

Groves, R., 1989: *Survey errors and survey costs*. New York.

Hershey, M./Hill, D., 1976: Positional response set in pre-adult socialisation surveys. In: *Social Science quarterly*, 56, 707-714.

Hippler, H.-J./Schwarz, N., 1992: The impact of administration mode on response effects in surveys. *ZUMA-Arbeitsberichte* 92/14. Mannheim, Germany.

Holaday, B./Turner-Henson, A., 1989: Response effects in surveys with school-age children. In: *Nursing Research*, 38, 248-250.

Knäuper, B., 1999: The impact of age and education on response order effects in attitude measurement. In: *Public Opinion Quarterly*, 63, 347-370.

Knäuper, B./Belli, R. Hill, D./Herzog, R., 1997: Question difficulty and respondent's cognitive ability: the effect on data quality. In: *Journal of Official Statistics*, 13, 181-199.

Kohlberg, L., 1976: Moral stages and moralization. The cognitive-developmental approach. In: Lickona, T., (Hrsg.), *Moral development and behavior. Theory, research and social issues*. New York, 31-53.

Krosnick, J., 1991: Response strategies for coping with the cognitive demands of attitude measures in surveys. In: *Applied Cognitive Psychology*, 5, 213-236.

Krosnick, J., 1992: The impact of cognitive sophistication and attitude importance on response order and question order effects. In: Schwarz, N./Sudman, S. (Hrsg.), *Context effects in social and psychological research*. New York, pp. 203-218.

Krosnick, J./Alwin, D., 1987: An evaluation of cognitive theory of response-order effect in survey measurement. In: *Public Opinion Quarterly*, 51, 201-219.

De Leeuw, E./Otter, M., 1995: The reliability of children's responses to questionnaire items: question effects in children's questionnaire data. In: Hox, J./van der Meulen, B./Kanssens, J./ter Laak, J./Tavecchio, L. (Hrsg.), *Advances in family research*. Amsterdam, 251-257.

Lyberg, L./Biemer, P./Collins, M./de Leeuw, E./Dippo, C./Schwarz, N./Trewin/D. (Hrsg.), 1997: *Survey measurement and process quality*. New York.

Marsh, H., 1986: Negative item bias in rating scales for preadolescent children: a cognitive-developmental phenomenon. In: *Developmental Psychology*, 22, 37-49.

Miller, W., 1976: Preschoolers' attitude toward libraries: can they be measured? In: *The Library Quarterly*, März 1976, 151-153.

Naplava, T./Oberwittler, D., 2002: Methodeneffekte bei der Messung selbstberichteter Delinquenz von männlichen Jugendlichen. In: *Monatsschrift für Kriminologie und Strafrechtsreform*, 85, 401-423.

Piaget, J., 1948: *The moral judgment of the child*. Glencoe, Ill.

Schuman, H./Presser, S., 1981: Questions and answers in attitudes surveys. Experiments on question form, wording, and context. Thousand Oaks.

Schwarz, N./Hippler, H.-J./Deutsch, B./Strack, F., 1985: Response scales. Effects of category range on reported behaviour and comparative judgements. In: *Public Opinion Quarterly*, 49, 388-395.

Schwarz, Norbert/Hippler, H.-J./Noelle-Neumann, E., 1992: A cognitive model of response order effects in survey measurement. In: Schwarz, N./Sudman, S. (Hrsg.), *Context effects in social and psychological research*. New York, 187-201.

Schwarz, N./Hipper, H.-J., 1995: The numeric values of rating scales: a comparison of their impact in mail surveys and telephone interviews. In: *International Journal of Public Opinion Research*, 7, 72-74.

Schwarz, N./Knäuper, B., 1999: Cognition, aging, and self-reports. In: Park, D./Schwarz, N. (Hrsg.), *Cognitive aging. A student primer*. Philadelphia, 233-252.

Schwarz, N./Knäuper, B./Hippler, H.-J./Noelle-Neumann, E./Clark/L., 1991a: Rating scales. Numeric values may change the meaning of scale labels. In: *Public Opinion Quarterly*, 55, 570-582.

Schwarz, N./Knäuper, B., Park, D., 1998: *Cognition, aging, and self-reports*. Philadelphia, PA.

Schwarz, N./Strack, F./Mai, H., 1991b: Assimilation and contrast effects in part-whole question sequences. A conversational logic analysis. In: *Public Opinion Quarterly*, 55, 3-23.

Scott, J., 1997: Children as respondents: Methods for improving data quality. In: Lyberg, L./Biemer, P./Coolins, M./de Leeuw, E./Dippo, C./Schwarz, N./Trewin, D. (Hrsg.), *Survey measurement and process quality*. Wiles, New York, 331-350.

Sudman, S./Bradburn, N./Schwarz, N., 1996: *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco.

Tourangeau, R./Rips, L./Rasinski, K., 2000: *The psychology of survey response*. New York.

Villancourt, P., 1973: Stability of children's survey responses. In: *Public Opinion Quarterly*, 37, 373-387.

Wänke, M./Schwarz, N., 1997: Reducing question order effects: The operation of buffer items. In: Lyberg, L./Biemer, P./Collins, M./de Leeuw, E./Dippo, C./Schwarz, N./Trewin, D., (Hrsg.), *Survey measurement and process quality*. New York, 115-140.

Korrespondenzadresse

PD Dr. Marek Fuchs

Universität Kassel

Fachbereich 5

D-34109 Kassel

email: marek.fuchs@uni-kassel.de

DETERMINANTEN UND KONSEQUENZEN DER UMFRAGEEINSTELLUNG

Bewertungsdimensionen unterschiedlicher Umfragesponsoren und die Antwortbereitschaft der Befragten

VOLKER STOCKÉ & BIRGIT BECKER ¹

Dieser Beitrag beschäftigt sich mit zwei Fragestellungen: a) den Bewertungen von Umfragen verschiedener Sponsoren auf den Dimensionen Nützlichkeit, Verlässlichkeit und Belastung als *Determinanten* der generalisierten Umfrageeinstellung, und b) der Beantwortung oder Verweigerung der Einkommensfrage als Indikator für die Kooperationsbereitschaft im Interview als *Konsequenz* der generalisierten Umfrageeinstellung. Im ersten Teil der Analyse wird auch die Bedeutung des quantitativen Ausmaßes der Interviewerfahrung für die Stärke der beobachteten Zusammenhänge berücksichtigt. Die empirische Analyse mit Daten einer lokalen Zufallsstichprobe zeigt zunehmend stärkere Zusammenhänge zwischen den sponsorenspezifischen Bewertungsdimensionen und der Umfrageeinstellung, wenn die Befragten bereits häufiger an Umfragen teilgenommen haben. Die wahrgenommene Nützlichkeit von Umfragen und die Bewertung wissenschaftlicher Sponsoren erweisen sich als die stärksten Bestimmungsfaktoren der generalisierten Umfrageeinstellung. Bezüglich der zweiten Fragestellung kann festgestellt werden, dass die Wahrscheinlichkeit einer Antwortverweigerung bei der Einkommensfrage sehr stark ansteigt, wenn die Befragten eine zunehmend negative und, gemessen an den Antwortlatenzen, zugleich kognitiv stark verankerte Umfrageeinstellung haben. Somit wird gezeigt, dass die generalisierte Umfrageeinstellung einen deutlichen Einfluss auf die Qualität von Umfragedaten hat.

1 Wir danken den anonymen Gutachtern sowie dem Herausgeber dieser Zeitschrift und ganz besonders Hartmut Esser für hilfreiche Kommentare und Anregungen. Eva Feuerbach und Diana Schirowski waren eine große Unterstützung bei der Manuskripterstellung. Der Beitrag wurde durch Förderung der Deutschen Forschungsgemeinschaft an den Sonderforschungsbereich 504 der Universität Mannheim ermöglicht.

This article deals with two questions: a) the evaluations of surveys of different sponsors on the dimensions utility, reliability and burden as *determinants* of the generalized attitude towards surveys, and b) the answer or refusal of the income question as an indicator of cooperative behavior during the interview as a *consequence* of respondents' attitudes towards surveys. In the first part of the analysis it is furthermore tested whether the quantity of survey experience in the past moderates the strength of the observed associations. The empirical analysis with data from a local survey based on a random probability sample shows increasingly stronger associations between respondents' sponsor-specific evaluations and their attitudes towards surveys when subjects have taken part more often in surveys in the past. The perceived utility of surveys and the evaluation of scientific sponsors proved to be the strongest determinants for the generalized attitude towards surveys. Regarding the second question of this article it is found that the probability of answering or refusing to answer the income question increases considerably when the interviewees have a more negative and - as indicated by their response latencies - at the same time cognitively accessible attitude towards surveys. Thus it is concluded that respondents' attitudes towards surveys have serious consequences for the quality of survey data.

1. Einleitung

Die Einstellung zu Umfragen ist eine wichtige Determinante für unterschiedliche Aspekte der Kooperationsbereitschaft von Befragten in sozialwissenschaftlichen Interviews. So haben Forscher festgestellt, dass sich die generalisierte Umfrageeinstellung auf die Elaboriertheit der Antworten auswirkt und einen Einfluss auf das Ausmaß des Item-Nonresponse während des Interviews hat (Reuband 1991; Rogelberg et al. 2001). Auch die Bereitschaft zur Interviewteilnahme sowie die Wiederbefragungsbereitschaft wird durch die Umfrageeinstellung beeinflusst (Erbslöh/Koch 1988; Schützenmeister 2002; Stinchcombe et al. 1981). Ob sie auch Einfluss auf die Wahrscheinlichkeit der Beantwortung sensibler Fragen hat, wie etwa der Frage nach dem Einkommen, wurde bisher nicht untersucht.

Wenn es sich, was die vorliegenden Ergebnisse nahe legen, bei der Umfrageeinstellung um eine wichtige Determinante der Kooperationsbereitschaft und Datenqualität handelt, so stellt sich die Frage nach deren Bestimmungsfaktoren. Es könnte vermutet werden, dass die Befragten ihre generalisierte Umfrageeinstellung auf ihre Bewertung konkreter Erscheinungsformen und spezieller Bewertungsdimensionen zurückführen. In dieser Hinsicht ist möglicherweise die Haltung der Probanden zu unterschiedlichen Typen von Auftraggebern von Umfragen bedeutsam. Einige ältere Studien haben zwar gezeigt, dass

sich die Sponsorschaft von Umfragen auf einige Aspekte der Kooperationsbereitschaft auswirkt (vgl. etwa: Doob et al. 1973; Peterson 1975), beim derzeitigen Forschungsstand ist jedoch ungeklärt, ob sich die differenzierte Bewertung verschiedener Sponsoren auch auf die Bewertung von Umfragen allgemein und damit potentiell auch auf die Kooperationsbereitschaft bei Befragungen durch ganz andere Auftraggebertypen auswirken. Der vorliegende Artikel zielt darauf ab, diese Erkenntnislücke zu schließen.

Dieser Artikel beschäftigt sich mit zwei Fragen: a) ob und wie stark die Bewertung von Umfragen unterschiedlicher Sponsoren eine *Determinante* der generalisierten Umfrageeinstellung ist, sowie b) ob die Beantwortung oder Verweigerung der Einkommensfrage als *Konsequenz* der generalisierten Umfrageeinstellung erklärt werden kann.

Im ersten Teil wird explorativ untersucht, wie die generalisierte Umfrageeinstellung durch die Haltung zu Befragungen unterschiedlicher Sponsoren und die hierbei möglicherweise bedeutsamen Bewertungsdimensionen beeinflusst wird. Konkret wird überprüft, ob sich die Umfrageeinstellung mehr durch die Einstellung gegenüber der Marktforschung, durch die Einstellung zu wissenschaftlichen Umfragen oder durch die Bewertung von Umfragen staatlicher Behörden vorhersagen lässt. Dabei wird danach differenziert, welche Bewertungskriterien für die Befragten dabei jeweils im Vordergrund stehen: Ist die wahrgenommene Nützlichkeit von Umfragen, deren angenommene Verlässlichkeit oder die erwartete Belastung durch Umfragen wichtiger? Unterscheidet sich der Einfluss der Bewertungsdimensionen nach den Sponsoren der Umfragen? Welche Bedeutung hat die Umfrageerfahrung als Determinante der Einstellungskristallisation für die Einstellungsstruktur?

Im zweiten Teil der Analyse wird überprüft, ob und wie sich die generalisierte Umfrageeinstellung auf die Beantwortung der, jedenfalls im Kontext deutscher Befragungen, sensiblen Frage nach dem Einkommen auswirkt. Dabei ist unsere Hypothese, dass Befragte mit einer positiven und gleichzeitig hinreichend kognitiv stark verankerten Bewertung von Umfragen eher zur Beantwortung der Einkommensfrage bereit sind.

2. Forschungsstand

2.1 Die generalisierte Einstellung zu Umfragen

In der Bundesrepublik wird die Teilnahme an Interviews von der Mehrheit der Befragten (ca. 50 Prozent) als angenehm empfunden, während nur ca. 15-20 Prozent der Probanden Befragungen als unangenehm bezeichnen. Diese Einschätzung scheint im Zeitverlauf relativ stabil zu sein (Forsa 1991, 1993a, 2000). Auch die Umfrageforschung insgesamt genießt ein gutes Image (Forsa 1993b).

Allerdings scheint die Einstellung zu Umfragen in Deutschland uneinheitlich zu sein. Leiblein und Klass (1990) konstatieren ein „recht ambivalentes Image“ der Umfrageforschung, was daran liegt, dass der Nutzen von Umfragen zwar eingesehen wird, aber Bedenken hinsichtlich des Datenschutzes und der willkürlichen Interpretation der Ergebnisse bestehen. Dieser Befund deckt sich auch mit den Ergebnissen von Porst, der insgesamt eine starke positive Grundhaltung zu Befragungen findet, jedoch auch eine gewisse Skepsis bezüglich der Seriosität und Anonymität bei Personen, die schon Erfahrung mit Umfragen haben. Auch sind manche Befragte der Meinung, aus Umfrageergebnissen könne „jeder genau das herauslesen, was ihm persönlich in den Kram passt“ (Porst 1998: 9).

In der amerikanischen Umfrageforschung werden steigende Zahlen von Umfrageskeptikern vermeldet. Die Anzahl der Personen, die eine negative Einstellung zu Umfragen äußern, ist zumindest in der Vergangenheit leicht angestiegen (Schleifer 1986). Dennoch wird meist festgehalten, dass von einer Krise der Markt- und Meinungsforschung keine Rede sein kann (Kohut 1986; McDaniel et al. 1985). Insgesamt ist die Einstellung zu Umfragen überwiegend positiv, wobei vor allem die Nützlichkeit von Umfragen von der Bevölkerung anerkannt wird. Zum Teil findet sich jedoch auch eine gewisse Skepsis, insbesondere im Bereich der Anonymität und der Verwendung der Daten.

Schließlich sollte angemerkt werden, dass Studien, die sich mit der Einstellung zu Umfragen befassen, wahrscheinlich einer Verzerrung dahingehend ausgesetzt sind, dass sie den Anteil von Personen mit positiver Einstellung in der Gesamtbevölkerung überschätzen. Schon die Teilnahme an einem Interview impliziert eine gewisse positive Haltung zu Umfragen, so dass nicht zu erwarten ist, dass die Ausfälle durch Verweigerung neutral sind. So sollte man das Ergebnis, dass die Skepsis gegenüber der Umfrageforschung relativ begrenzt ist und die positiven Aspekte überwiegen, unter dem Vorbehalt sehen, dass sich die größten Umfrageskeptiker möglicherweise erst gar nicht dazu befragen lassen.

2.2 Determinanten der Umfrageeinstellung: Die Rolle der Sponsorenschaft

Eine mögliche Determinante der Umfrageeinstellung ist die Sponsorenschaft von Umfragen. Es besteht die Möglichkeit, dass die Befragten zwischen Auftraggebern von Umfragen und deren Bewertung differenzieren. Studien, die sich direkt mit dem Einfluss der Sponsoren auf die Umfrageeinstellung beschäftigen, existieren bisher nicht. Jedoch finden sich Hinweise darauf, dass sich die Sponsorenschaft auf die Einstellung zum jeweiligen Interview und dadurch vermittelt auf die Kooperationsbereitschaft im konkreten Umfragekontext auswirkt. Die beobachteten Zusammenhänge zwischen der Teilnahmebereitschaft, dem Ausmaß des Item-Nonresponse, der Korrektheit der Angaben einerseits und

der Sponsorenschaft der Umfrage andererseits können als indirekte Evidenz für sponsorenspezifische Bewertungen von Umfragen angesehen werden. Es kann vermutet werden, dass die Befragten ihre unterschiedlichen Urteile über Befragungen durch verschiedene Sponsoren auf ihre Bewertung von Umfragen insgesamt generalisieren.

Die meisten empirischen Studien beschäftigen sich mit Unterschieden in der Kooperationsbereitschaft der Befragten zwischen universitären und kommerziellen Sponsoren. Doob und Kollegen (1973) haben in ihrer postalischen Befragung den Sponsor (Stanford University vs. Industrial Research Associates) sowie die Teilnahmeanreize (0 vs. 5 c vs. 20 c) variiert. Sie kommen zu dem Ergebnis, dass in der Bedingung ohne Anreize die Rücklaufquote bei dem universitärem Sponsor signifikant höher ist als bei dem kommerziellem Sponsor, bei höheren Anreizen aber kein Unterschied vorliegt. In der Studie von Schöne (1973) wird der Auftragebereffekt ebenfalls durch den Rücklauf bei einer postalische Befragung gemessen. Zwar wird hier ein höherer Rücklauf bei einem universitärem Sponsor (Sozialpsychologisches Institut der Universität zu Köln) im Vergleich zu einem kommerziellem Sponsor (Institut für Markt- und Werbeforschung Köln, GmbH und Co. KG) verzeichnet, der Unterschied ist jedoch insignifikant. Peterson (1975) berichtet beim Vergleich zwischen einer Universität (University of Texas) und einem Marktforschungsinstitut (Group Seven Associates) als Auftraggeber einen signifikant höheren Rücklauf des schriftlichen Fragebogens und weniger Item-Nonresponse, wenn die Universität als Auftraggeber genannt wurde. Dagegen antworten die Befragten signifikant schneller beim Marktforschungsinstitut.

In einem Experiment vergleicht Orr (1976) die Auswirkungen der Sponsorenschaft auf die wahrgenommene Legitimität der Interviewanfrage, die Unterstützung des Sponsors sowie die Beantwortung intimer Fragen (Nonresponse und Offenheit der Antworten). Als Sponsor wurde einmal die „Holbrooke Foundation“ genannt, die im Einleitungsschreiben als „non-profit“-Institution im Wohlfahrtsbereich beschrieben wurde. In der anderen Bedingung wurde als Sponsor die „Holbrooke Marketing Research Inc.“ vorgestellt. Der Autor kann nachweisen, dass die Probanden einem „non-profit“-sponsor“ eine signifikant größere Legitimität zubilligen und weitere Forschung dieses Sponsors stärker befürworten als bei einem „profit-making“-sponsor. Auch die Bereitschaft zur Beantwortung sensibler Fragen und die Offenheit dieser Antworten (beurteilt durch drei Experten) ist in der Bedingung des „non-profit“-Sponsors signifikant höher.

Bei einem Vergleich der Teilnahmebereitschaft zwischen einer Universität (Bureau of Business Research of the University of Oregon), einem fiktiven Marktforschungsinstitut (Attitude Research Associates) und einem Kaufhaus findet Hawkins (1979) bei einer postalischen Befragung die höchste Rücklaufquote für die Universität, die sich jedoch

nicht signifikant von der Marktforschungs-Bedingung unterscheidet, beide sind aber signifikant höher als bei der Kaufhaus-Bedingung.² Jones und Lang (1983) finden bei einer postalischen Befragung einen höheren Rücklauf und weniger Item-Nonresponse, wenn eine Universität statt einer privaten Marktforschungsfirma als Sponsor der Umfrage identifiziert werden konnte.

Es gibt jedoch auch Studien, deren Ergebnisse bezüglich der Sponsorenschaft weniger eindeutig ausfallen. Houston und Nevin (1977) können in ihrer Studie keinen signifikanten Haupteffekt zwischen einem universitärem (Graduate School of Business, University of Wisconsin-Madison) und einem kommerziellen Sponsor (Business Research Associates) bezüglich der Rücklaufquote eines postalischen Fragebogens finden; jedoch zeigt sich bei dem universitären Sponsor signifikant weniger Item-Nonresponse. Sie finden zudem eine signifikante Interaktion zwischen dem Sponsor und der Art des Appells (altruistisch vs. egoistisch) auf die Rücklaufquote, wobei ein universitärer Sponsor mit altruistischen Appellen am erfolgreichsten ist.

Albaum (1987) betont, dass man bei der Untersuchung von Sponsoreffekten darauf achten sollte, zwischen dem Auftraggeber (sponsor) und der durchführenden Organisation (source) einer Umfrage zu unterscheiden. In seiner Studie hat Albaum den Durchführenden variiert (Universität vs. privates Umfrageinstitut vs. Kreditgenossenschaft), während als Auftraggeber immer die Kreditgenossenschaft ersichtlich war. Befragt wurden Mitglieder dieser Kreditgenossenschaft mittels eines schriftlichen Fragebogens. In der Universitäts-Bedingung konnte der höchste Rücklauf erzielt werden, allerdings waren alle Unterschiede zwischen den verschiedenen Source-Bedingungen nicht signifikant. Bezüglich der Antwortgeschwindigkeit schnitt die Universität signifikant besser ab als die beiden anderen durchführenden Organisationen. Der Anteil ausgelassener Items war bei der Nennung der Kreditgenossenschaft als durchführende Institution höher als in den beiden anderen Bedingungen, allerdings waren die Unterschiede hier nicht signifikant.

In einer Untersuchung von Rogelberg und Kollegen (2003) hat sich gezeigt, dass die Bewertung des Sponsors einer Umfrage nur für die Teilgruppe der Personen mit geringer Teilnahmebereitschaft eine Rolle spielt. In dieser Studie war für alle Mitglieder der

2 In dieser Studie war die Variation des Sponsors mit der Angabe über die Verwendung der Daten konfundiert. Während in der Universitäts-Bedingung der Satz „Furthermore, your responses will not be given to any business firm“ hinzugefügt war, erschien in der Kaufhaus-Bedingung die Aussage „The results will be used by the store’s management to develop more effective marketing strategies in order to better serve the Eugene community“ (Hawkins 1979: 581). Somit ist fraglich, ob die gefundenen Effekte auf die Sponsorenschaft und nicht auf unterschiedlich starke Datenschutzbedenken zurückgehen.

Grundgesamtheit die Zufriedenheit mit dem Sponsor und die Wiederbefragungsbereitschaft aus einer vorherigen Befragung bekannt. Die Teilnahmebereitschaft bei einer nachfolgenden postalischen Umfrage hat sich nur bei Befragten mit geringer Wiederbefragungsbereitschaft nach der Bewertung der Sponsorenorganisation unterschieden.

Insgesamt lässt sich aus den dargestellten Befunden ableiten, dass die Befragten wissenschaftliche Sponsoren eher unterstützen als kommerzielle Institute. Allerdings kann dieser Befund auch in einer anderen Art interpretiert werden. Presser und Kollegen (1992) sehen den Grund für die höhere Kooperationsbereitschaft im Falle universitärer Sponsoren in erster Linie in deren höheren Bekanntheit (vor allem im Vergleich zu relativ unbekanntem oder gar fiktiven Marktforschungsinstituten). Daher variierten sie in ihrer Studie den Sponsor zwischen einer bekannten Universität (University of Maryland) und einer bekannten Zeitung (Washington Post). Die Teilnahmebereitschaft war beim universitären Sponsor etwas höher, jedoch nicht statistisch signifikant. Damit sehen die Autoren ihre Hypothese prinzipiell bestätigt, räumen aber ein, dass beide Sponsoren als „im Interesse der Bevölkerung“ wahrgenommen werden könnten und dieses Kriterium möglicherweise für Sponsoreffekte verantwortlich ist.

2.3 Konsequenzen der generalisierten Umfrageeinstellung: Kooperationsbereitschaft

Zahlreiche empirische Untersuchungen zeigen, dass die generalisierte Einstellung zu Umfragen eine wichtige Determinante für die Kooperationsbereitschaft der Befragten im Befragungskontext und damit ein zentraler Bestimmungsfaktor der Qualität von Umfragedaten ist. Personen mit einer positiven Umfrageeinstellung nehmen eher an einer Umfrage teil als Personen mit einer negativen Haltung. Dieser Zusammenhang zwischen Umfrageeinstellung und Teilnahmebereitschaft konnte in einer Reihe empirischer Untersuchungen nachgewiesen werden (vgl. Erbslöh/Koch 1988; Goyder 1986; Sharp/Frankel 1983; Stinchcombe et al. 1981). So ist der Rekrutierungsaufwand, der bei Personen mit negativer Umfrageeinstellung benötigt wird, höher als bei positiv eingestellten Personen. Es zeigt sich, dass Verweigerer, die zuerst eine Befragung abgelehnt haben, dann aber doch noch zu einem Interview bewegt werden konnten, wesentlich skeptischer und negativer zu Umfragen eingestellt sind als spontane Teilnehmer (Erbslöh/Koch 1988; Stinchcombe et al. 1981). Befragte mit einer positiven Umfrageeinstellung haben eine höhere Wiederbefragungsbereitschaft (Rogelberg et al. 2001; Schützenmeister 2002). Auch die Verweigerung einzelner Fragen während des Interviews wird von der Umfrageeinstellung beeinflusst: Eine negative Umfrageeinstellung korrespondiert mit höherem Item-Nonresponse (Rogelberg et al. 2001). Weitere Belege für den Zusammenhang zwischen der Umfrageeinstellung und der Kooperationsbereitschaft ergeben sich daraus, dass

Teilnehmer einer postalischen Befragung mit einer positiven Umfrageeinstellung sich stärker an die Anweisungen des Fragebogens halten (z.B. nicht mehrere Antworten ankreuzen, wenn nur eine erlaubt ist) und die Fragebögen schneller ausgefüllt zurückschicken (Rogelberg et al. 2001). Schließlich gibt es Hinweise, dass auch die Validität von Befragtenantworten von deren Umfrageeinstellung abhängt. So wurde in einer Studie von Stocké (2004) die generalisierte Einstellung zu Umfragen erfasst und überprüft, ob diese vorhersagt, in welchem Umfang die Angaben der Befragten über ihre Ausländereinstellung durch Anreize zu sozial erwünschtem Antwortverhalten beeinflusst werden. Diese Anreize wurden durch einen multiplikativen Index der von den Befragten subjektiv wahrgenommenen Erwünschtheit unterschiedlicher Einstellungsantworten, durch deren Bedürfnis nach sozialer Anerkennung sowie der Privatheit der Antwortsituation operationalisiert. Es hat sich herausgestellt, dass Befragte mit einer positiven und zugleich, nach der Geschwindigkeit der Beantwortung der Umfrageteilnahme, kognitiv stark zugänglichen Bewertung von Umfragen, vollständig durch Antwortverzerrungen durch soziale Erwünschtheit unbeeinflusst geblieben sind. Für alle anderen Befragten wurde eine starke Empfänglichkeit für Anreize durch soziale Erwünschtheit festgestellt. Insgesamt zeigt sich in diesen Ergebnissen, dass nicht nur die Teilnahmequote, sondern auch die in den Interviews realisierte Datenqualität von der Umfrageeinstellung der Probanden abhängt.

Es gibt jedoch schon lange Zweifel daran, ob eine bestimmte Einstellung immer zum entsprechenden Verhalten führt. In der Einstellungsforschung hat es kontroverse Debatten darüber gegeben, ob bzw. unter welchen Umständen Einstellung und Verhalten konsistent sind (vgl. Six/Eckes 1996). Diese Bedingungen sind dafür relevant, wann genau die Umfrageeinstellung die Kooperationsbereitschaft der Befragten und damit die Qualität von Umfragedaten beeinflusst.

Eine besondere Bedeutung in der Diskussion um die Einstellungs-Verhaltenskonsistenz nimmt die Einstellungsstärke ein. Verschiedene Teildimensionen dieses Konstruktes, wie z.B. die Einstellungszugänglichkeit, haben sich hierbei als signifikante Moderator-Variablen erwiesen (Bassili 1993, 1995; Fletcher 2000; Kokkinaki/Lunt 1997). Die kognitive Zugänglichkeit von Einstellungen wird meist durch Antwortlatenzen operationalisiert, d.h. die Geschwindigkeit der Antwortgenerierung wird als Indikator für die Abrufbarkeit der Einstellung aus dem Gedächtnis angesehen.

Studien, die sich mit der Moderatorfunktion von Antwortlatenzen für die Konsistenz zwischen Einstellung und Verhalten beschäftigt haben, kommen überwiegend zu bestätigenden Ergebnissen. In zahlreichen Untersuchungen wurde nachgewiesen, dass der Zusammenhang zwischen den Einstellungen gegenüber politischen Kandidaten und dem tatsächlichen Wahlverhalten größer ist, wenn die Befragten auf die Einstellungsfrage

schnell geantwortet haben (Bassili 1993, 1995; Bassili/Bors 1997; Fazio/Williams 1986; Fletcher 2000). Auch beim Themenkomplex Konsumprodukte gibt es Evidenzen für die Moderatorrolle der Antwortgeschwindigkeit. Kokkinaki und Lunt (1997) finden eine zunehmende Übereinstimmung zwischen den Einstellungen zu Produkten und der Konsumabsicht, wenn die Einstellungsfragen zunehmend schneller beantwortet werden. Zu dem gleichen Ergebnis kommen Fazio, Powell und Williams (1989) in einem Experiment, in dem die Versuchspersonen ihre Einstellungen zu verschiedenen Süßigkeiten durch Drücken einer „like“- oder „dislike“-Taste zum Ausdruck brachten. Auch hier wurden die Reaktionszeiten als Indikator für die Verfügbarkeit der Einstellungen verwendet. Als Belohnung für die Teilnahme am Experiment durften sich die Versuchspersonen fünf der Produkte aussuchen. Die Korrelation zwischen der Einstellung gegenüber einem Produkt und dessen Auswahl war eine lineare Funktion der Einstellungsverfügbarkeit: Je schneller die Versuchspersonen auf die Einstellungsfrage antworteten, desto wahrscheinlicher war die Produktwahl konsistent mit dieser Einstellung.

Für die konkrete Vorhersage des Zusammenhangs zwischen der generellen Umfrageeinstellung und der Beantwortung der Frage nach dem Einkommen (als Indikator für kooperatives Verhalten) kann das MODE-Modell von Fazio (1990) herangezogen werden. In diesem Modell sind die spontane Aktivierbarkeit der Einstellungsurteile sowie die Elaboriertheit der Informationsverarbeitung die Determinanten für das Ausmaß an Einstellungsverhaltenskonsistenz. Im Falle einer elaborierten Informationsverarbeitung generieren die Akteure in der Handlungssituation *spezifische* Bewertungen zu dem konkreten Verhalten und es ist eine hohe Konsistenz zwischen diesen spezifischen Bewertungen und dem Verhalten erwartbar. Dagegen ziehen die Akteure ihre *generalisierten* Einstellungen als Entscheidungsgrundlage heran, wenn sie sich in einem automatischen Informationsverarbeitungsmodus befinden. Mit einem solchen Verarbeitungsmodus muss dann gerechnet werden, wenn die Akteure entweder nicht hinreichend zu einer aufwendigeren, elaborierten Art der Handlungsselektion motiviert sind oder hierzu, etwa wegen hohem Zeitdruck, die Gelegenheit fehlt. Ob unter diesen Umständen eine hohe Übereinstimmung zwischen den Einstellungsurteilen und dem Verhalten zu erwarten ist, hängt zusätzlich von der Stärke der kognitiven Zugänglichkeit dieser Urteile ab: Je zugänglicher die Einstellung ist, desto größer ist deren Prägekraft für das Handeln.

Es muss davon ausgegangen werden, dass ein Interview typischerweise eine Handlungssituation mit geringem Konsequenzenpotenzial, also eine „low cost“-Situation darstellt. In diesem Situationstyp ist die Motivation der Befragten für aufwendige Entscheidungsfindungen begrenzt. Für solche Situationen wird nach dem MODE-Modell eine spontane Art der Informationsverarbeitung vorhergesagt, unter der die Einstellung der Befragten zu Umfragen das (Antwort-) Verhalten beeinflussen kann. Ob dies jedoch geschieht, hängt in

diesem konkreten Anwendungsfall von der kognitiven Verfügbarkeit dieser Einstellung ab. Es kann argumentiert werden, dass es sich bei Umfragen für die durchschnittlichen Befragten um einen wenig salienten Einstellungsgegenstand handelt, so dass mit einer durchschnittlich geringen Einstellungsstärke und damit Einstellungszugänglichkeit gerechnet werden muss. Demnach müssen bei der Analyse der Einstellungsverhaltenskonsistenz in diesem Bereich Unterschiede in der Einstellungszugänglichkeit zwischen den Befragten als besonders erklärungskräftig angesehen werden.

Zu der gleichen Vorhersage kommt das Modell der Frame-Selektion von Esser, das eine differenziertere Modellierung zu Grunde legt (Esser 2001: 259 ff.). In einer konkreten Entscheidungssituation findet eine zweifache Selektion statt: die Modell-Selektion, die aus der Auswahl eines gedanklichen Modells der Orientierung besteht, sowie die Modus-Selektion, die die Art der Informationsverarbeitung bestimmt. Auch im Rahmen dieser Theorie wird in einer Befragungssituation eine spontane Art der Informationsverarbeitung vorausgesagt, da sich eine aufwendige Informationssuche und Reflexion aus der Sicht der Befragten kaum lohnen dürfte. Ob nun das Modell „kooperativer Befragter“ gewählt wird oder nicht, ist zum einen vom Nutzen, zum anderen von der Modellgeltung dieses Frames abhängig. Der Nutzen eines kooperativen Verhaltens in einer Umfragesituation ergibt sich beispielsweise daraus, dass der Befragte an einer akkurat durchgeführten Umfrage und der Verwendbarkeit der Ergebnisse für die Gesellschaft interessiert ist. Eine Teildimension der Modellgeltung ist die Zugänglichkeit des gedanklichen Modells in der Identität des Akteurs (Esser 2001: 270). Im konkreten Anwendungsfall kann angenommen werden, dass das Modell „kooperativer Befragter“ dann besonders zugänglich ist, wenn eine positive Umfrageeinstellung vorhanden und diese gleichzeitig kognitiv stark verfügbar ist.

3. Datengrundlage und Operationalisierung

3.1 Stichprobe und Repräsentativität der Daten

Die Studienteilnehmer entstammen einer lokalen, mehrstufigen Zufallsstichprobe der Wohnbevölkerung einer süddeutschen Großstadt. Die Grundgesamtheit der im Sommer 1999 durchgeführten Untersuchung bildeten Personen mit Lebensmittelpunkt im Stadtgebiet, die zum Befragungszeitpunkt mindestens 18 Jahre alt waren und die deutsche Staatsbürgerschaft besaßen. Mit einer Zufallsbegehung wurde durch die Mitarbeiter des Forschungsprojektes eine Haushaltsstichprobe gezogen und den Interviewern vorgegeben (Adress-Random). Die Startpunkte der Begehung wurden nach dem Zufallsprinzip im Stadtgebiet bestimmt. Innerhalb der Haushalte wurde immer die Person für eine Befragung ausgewählt, die vor dem Kontaktzeitpunkt als letzte Geburtstag hatte. Mit 139 verwertbaren Interviews ließ sich eine Ausschöpfungsquote von 40.4 Prozent realisieren. Ein

Vergleich der Zusammensetzung der in der Nettostichprobe repräsentierten Befragten mit den teilweise bekannten und teilweise durch Daten auf höherer Gebietsebene angenäher-ten soziostrukturellen Merkmale des Auswahlgebietes ergibt die üblicherweise in Umfragedaten vorliegenden Abweichungen.³

3.2 Vorgehensweise

Die Befragung erfolgte in allen Fällen als computergestütztes Interview bei den Befragten zuhause. Die Umfrage wurde den Befragten in einem vorab zugestellten Anschreiben als Befragung über „aktuelle politische Themen“ und über „die Bewertung von und Erfahrung mit Umfragen“ angekündigt. Im postalisch zugestellten Anschreiben wurde angegeben, dass es sich um eine von der lokal ansässigen Universität durchgeführte Befragung der Bürger im Stadtgebiet handelt. Die Interviews dauerten im Durchschnitt 35 Minuten.

3.3 Operationalisierung

Unser Untersuchungsziel erfordert die Operationalisierung der folgenden Faktoren:

- *Generalisierte Umfrageeinstellung*: Auf der Grundlage vorliegender Studien (Dran/Hildreth 1995; Erbslöh/Koch 1988; Forsa 1993a, 1993b, 1996; Goyder 1986; Porst 1998; Stinchcombe et al. 1981) wurden 16 Items zur Erfassung der generalisierten Einstellung zu Umfragen ausgewählt (vgl. die Itemformulierungen in Tabelle 1 unten). Diese 16 Einzelitems bekamen die Befragten zu Beginn des Interviews in randomisierter Form dargeboten. Die Zustimmung der Befragten zu den positiven und negativen Iteminhalten wurde auf einer siebenstufigen Antwortskala (1 „stimme voll und ganz zu“ bis 7 „lehne voll und ganz ab“) erhoben. Nach den Ergebnissen einer Hauptkomponentenanalyse mit Varimax Rotation der Faktoren besteht die Gesamtskala aus drei Teilskalen. Die formale Gültigkeit der Gesamtskala ist daher nach Cronbachs Alpha mit einem Wert von .73 relativ schwach.

3 In der Stichprobe sind Personen mit Abitur um 11, verheiratete Befragte um 10 und Angestellte um 18 Prozentpunkte überrepräsentiert. Dies führt zu einer Unterrepräsentanz der jeweils anderen Bildungs-, Familien- und Erwerbsstatusgruppen. Beim Einkommen der Befragten tritt der bei Umfragen bekannte Mittelschichtbias auf, so dass in der Stichprobe Personen mit geringem und hohem Einkommen unterrepräsentiert sind. Dagegen liegen hinsichtlich des Alters und des Geschlechtes der Befragten keine systematischen Abweichungen von der Zusammensetzung der Grundgesamtheit vor. Die Vergleichsdaten für die Bildungs- und Ausbildungsabschlüsse, den Familienstand und die Einkommensverteilung beziehen sich mangels spezifischer Daten auf die Bevölkerung Westdeutschlands im Dezember 1996. Die Aussagen über die Repräsentativität der im Datensatz vertretenen Geschlechts- und Altersstruktur beruhen auf spezifischen Zahlen der Stadt, welche die Grundgesamtheit definiert (Stadt Mannheim 1998).

Durch die Verkürzung der Skala um 4 Items lässt sich eine eindimensionale Faktorenstruktur erreichen und die Reliabilität auf ein Alpha von .80 erhöhen. Unsere Analysen haben aber gezeigt, dass eine solche Verbesserung der formalen Gültigkeit zulasten der inhaltlichen Gültigkeit des Instrumentes geht. Der vollständigen Skala kommt nach den in Abschnitt 4.3 berichteten Ergebnissen statistisch signifikante Prognosekraft für die Antwortbereitschaft der Befragten zu. Die Konstruktvalidität wird durch eine Reduktion der Itemanzahl soweit geschwächt, dass eine Vorhersage der Antwortverweigerungen nur noch eingeschränkt und auf einem statistisch nicht mehr abgesicherten Niveau möglich ist. Aus diesem Grund haben wir uns trotz der eingeschränkten formalen Gültigkeit zur Verwendung der Skala in vollem Umfang entschieden. Nach der Angleichung der Item-Polung wurde für jede befragte Person ein Durchschnittswert ihrer Einstellungsangaben über Umfragen berechnet. Die resultierende Skala variiert zwischen dem Wert 1 (sehr negative Einstellung) und dem Wert 7 (sehr positive Einstellung).

- *Sponsorenspezifische Bewertungsdimensionen:* Den Befragten wurden im Anschluss an die allgemeinen Fragen über Umfragen 9 Items vorgelegt, die jeweils die Nützlichkeit, Verlässlichkeit und Belastung für die Sponsoren Universität, staatliche Behörde und kommerzielles Marktforschungsinstitut erfassen. Diese Sponsoren wurden ausgewählt, da aufgrund der bisherigen Forschung wahrscheinlich ist, dass Befragte gerade zwischen diesen differenzieren. Die drei Bewertungsdimensionen wurden gemessen, indem für jeden Sponsor gefragt wurde, ob Umfragen dieses Sponsors als nützlich angesehen werden (Nutzen-Dimension), ob die Ergebnisse dieses Sponsors zuverlässig sind (Verlässlichkeits-Dimension) und ob die Teilnahme an Umfragen dieses Sponsors unangenehm ist (Belastungs-Dimension) (vgl. die Formulierung der Frage und der Items im Anhang). Die Interpretation dessen, was an Umfragen nützlich oder belastend sein kann und worin die Bestimmungsfaktoren deren Verlässlichkeit liegen, wurde den Befragten überlassen. Die Zustimmung zu diesen Items wurden auf einer siebenstufigen Antwortskala mit den Endpunkten 1, „lehne voll und ganz ab“ und 7, „stimme voll und ganz zu“ erfasst.

- *Kognitive Zugänglichkeit der Umfrageeinstellung:* Als Indikator der kognitiven Zugänglichkeit der Umfrageeinstellung wird die durchschnittliche Geschwindigkeit der Beantwortung der 16 Umfrageitems verwendet. Die von den Befragten benötigte Zeit wurde zusammen mit den inhaltlichen Angaben durch die Interviewer im Rahmen des computer-administrierten Fragebogens erfasst. Die durch die Software bedingte technische Obergrenze der Messgenauigkeit liegt bei einem Hundertstel einer Sekunde.⁴

4 Messstörungen, etwa durch Ablenkung der Probanden oder Verständnisfragen, haben zu 9.3% invaliden Latenzmessungen geführt. Da angenommen werden kann, dass diese Ausfälle nicht zu-

- *Umfrageerfahrung*: Als Indikator für das vorliegende Ausmaß an Interviewerfahrung haben die Befragten einerseits angegeben, ob sie vor der aktuellen Befragung bereits an einer anderen Umfrage teilgenommen haben. Andererseits wurde erfasst, wie häufig dies insgesamt, abgesehen von dem aktuellen Interview, bereits der Fall war.

- *Antwortbereitschaft*: Die kategorische Antwortverweigerung bei der im Rahmen einer deutschen Umfrage besonders „heiklen“ Frage nach dem monatlichen Netto-Haushaltseinkommen wurde als Indikator für die Kooperationsbereitschaft der Befragten herangezogen. Die Frage nach dem Haushaltseinkommen wurde am Ende des soziodemographischen Blocks gestellt, der sich an den Themenblock „Einstellungen zu Umfragen“ anschloss. Bei der zweistufigen Vorgehensweise wurde zuerst bei allen Probanden offen nach dem Haushaltseinkommen gefragt. Haben die Befragten hierbei eine Angabe verweigert oder angegeben, das Haushaltseinkommen nicht zu kennen, so wurde in einem zweiten Schritt eine Erfassung von Einkommenskategorien angestrebt. Zu dem Zweck wurde den Befragten eine Antwortskala mit 22 Kategorien vorgelegt und gefragt, ob sie zumindest eine der Einkommenskategorien nennen wollen oder können.⁵ Als Indikator für die Bereitschaft der Befragten zu Beantwortung „heikler“ Fragen und damit für das Ausmaß ihrer Kooperationsbereitschaft wurden die Befragten danach aufgeteilt, ob sie einerseits eine Angabe über ihre Haushaltseinkommen gemacht haben (wenn auch eventuell nur in der Form einer Einkommenskategorie) bzw. angegeben haben, den Betrag nicht zu kennen oder andererseits eine Antwort verweigert haben.

4. Ergebnisse

Die empirischen Ergebnisse unserer Studie werden in drei Abschnitten dargestellt. Der *erste* Teil ist eine deskriptive Darstellung der generalisierten Umfrageeinstellung und der ansonsten verwendeten Variablen. Im *zweiten* Abschnitt wird explorativ die Erklärungskraft der sponsorenspezifischen Bewertungsdimensionen und die der Umfrageerfahrung für die Einstellung zu Umfragen untersucht. Im *dritten* Teil werden dann die in Abschnitt 2.3 theoretisch abgeleiteten Vorhersagen über die Konsequenzen der Umfrageeinstellung für die Bereitschaft der Befragten zur Beantwortung der Einkommensfrage überprüft.

fällig erfolgt sind und eine systematische Stichprobenselektion nach den Störungsursachen vermieden werden soll, wurden die fehlenden Werte durch eine Mittelwert-Imputation ergänzt.

5 Der Fragetext lautet wie folgt: „Ich kann Ihnen zusichern, das Ihre Angaben absolut anonym und vertraulich behandelt werden. Vielleicht können Sie mir zumindest sagen, in welcher Größenordnung sich Ihr Nettoeinkommen bewegt. Geben Sie mir einfach die entsprechende Nummer auf der vorliegenden Liste an.“

4.1 Deskriptive Ergebnisse

Es zeigt sich, dass die Befragten insgesamt eine positive Einstellung zu Umfragen äußern (vgl. Tabelle 1). Der Gesamtmittelwert der generalisierten Umfrageeinstellung liegt bei 4.66 auf der sieben-stufigen Antwortskala. Trotz der anzunehmenden, im Vergleich zur Grundgesamtheit positiven Selbstselektion der Studienteilnehmer zeigt sich eine hohe Variabilität der Bewertung von Umfragen: Die individuellen Skalenwerte bewegen sich zwischen 2.81 und 6.75. Auch hinsichtlich der Einzelitems gibt es sehr unterschiedliche Einstellungsäußerungen. Während etwa die Wichtigkeit von Umfragen für Wissenschaft, Politik und Wirtschaft stark anerkannt wird (Mittelwert 5.44), werden die bei einer Umfrageteilnahme notwendigen Anstrengungen eher kritisch gesehen (Mittelwert 3.89).

Tabelle 1: Items zur Erfassung der generalisierten Umfrageeinstellung

Itemformulierung	Antworten	Antwort-
	Ø (Std.)	latenz Ø (Std.)
1. Umfragen sind wichtig für Wissenschaft, Politik und Wirtschaft	5.44 (1.56)	2.97 (4.15)
2. Meinungsumfragen machen die Gesellschaft demokratischer	4.37 (1.83)	4.97 (5.95)
3. Die Teilnahme an Umfragen ist in meinem eigenen Interesse	4.60 (1.83)	4.23 (5.30)
4. Es sollte für jeden selbstverständlich sein, an einer Umfrage teilzunehmen	4.66 (1.97)	4.26 (7.29)
5. Umfragen bringen Abwechslung und sind interessant	4.50 (1.73)	3.33 (5.16)
6. Bei Umfragen kann man endlich mal seine Meinung sagen	4.89 (1.92)	4.64 (7.75)
7. Die Ergebnisse von Umfragen sind in den meisten Fällen richtig	4.73 (1.53)	5.57 (8.12)
8. Die meisten Umfragen sind seriös und werden verantwortungsvoll durchgeführt	4.70 (1.53)	4.94 (8.79)
9. Bei Umfragen werden normale und repräsentative Menschen befragt	4.86 (1.83)	5.62 (6.95)
10. Die Teilnehmer an Umfragen sagen meistens ihre wahre Meinung	5.01 (1.74)	4.08 (4.50)
11. Die Teilnehmer an Umfragen geben sich sehr viel Mühe, richtig zu antworten	5.30 (1.47)	4.60 (5.36)
12. Durch Umfragen wird man nur von wichtigeren Dingen abgehalten	4.70 (1.81)	4.85 (6.90)
13. Bei Umfragen wird häufig etwas gefragt, was niemanden etwas angeht	4.02 (1.87)	6.31 (7.60)
14. In Deutschland werden viel zu viele Umfragen durchgeführt	4.17 (1.90)	4.07 (7.16)
15. Es ist anstrengend in Umfragen so viele Fragen zu beantworten	3.89 (2.01)	4.88 (6.64)
16. Interviewer sind unangenehme und aufdringliche Leute	4.77 (1.92)	5.78 (8.93)
Gesamtskala	4.66 (0.79)	4.69 (3.48)
Fallzahl	N=139	

Antwortskala: Wert 1 = negative Umfrageeinstellung; Wert 7 = positive Umfrageeinstellung. Die Angaben bei den Items 1 bis 11 wurden dieser Polung angepasst. Die Maßeinheit der Antwortlatenzen ist in Sekunden.

Die für die Beantwortung der Fragen nach der Umfrageeinstellung durchschnittlich benötigte Zeit variiert zwischen 2.97 Sekunden bei Item 1 und 6.31 Sekunden bei Item 13. Im Mittel haben die Befragten 4.69 Sekunden zur Generierung ihrer Antworten über die Umfrageeinstellung benötigt.

Bezüglich der sponsorenspezifischen Umfrageeinstellung zeigt sich eine klare Bewertungshierarchie (vgl. Tabelle 2): Wissenschaftliche Befragungen haben die positivste (Mittelwert 5.03) und kommerzielle Umfragen die negativste Reputation (Mittelwert 4.39). Befragungen durch staatliche Institutionen sind dazwischen angesiedelt (Mittelwert 4.81). Alle Bewertungen der unterschiedlichen Sponsoren sind nach zweiseitigen t-Tests signifikant verschieden (wissenschaftlich vs. kommerziell: $t=5.9$, $df=138$; staatlich vs. kommerziell: $t=4.1$, $df=138$; wissenschaftlich vs. staatlich: $t=2.6$, $df=138$; alle Tests: $p < 0.05$).

Tabelle 2: Bewertung von Umfragen unterschiedlicher Sponsoren hinsichtlich ihrer Nützlichkeit, Verlässlichkeit und der Belastung bei einer Teilnahme

	Kommerzielle Umfragen	Wissenschaftliche Umfragen	Staatliche Umfragen	Gesamt
	Ø (Std.)	Ø (Std.)	Ø (Std.)	Ø (Std.)
Nützlichkeit	4.80 (1.58)	5.50 (1.53)	5.22 (1.64)	5.17 (1.15)
Verlässlichkeit	4.25 (1.61)	4.78 (1.46)	4.60 (1.55)	4.54 (1.19)
Belastung ^{R)}	4.14 (1.85)	4.81 (1.86)	4.44 (1.79)	4.46 (1.37)
Gesamt	4.39 (1.13)	5.03 (1.03)	4.81 (0.94)	4.81 (0.94)

Skalenwert 1 = negative Umfrageeinstellung; Skalenwert 7 = positive Umfrageeinstellung. Die Angaben bei den mit ^{R)} markierten Items wurden an die Polung der anderen Angaben angepasst.

Werden Umfragen auch nach den drei Dimensionen Nützlichkeit, Verlässlichkeit und wahrgenommene Belastungen unterschiedlich bewertet? Es zeigt sich, dass Umfragen bezüglich ihrer Nützlichkeit mit einem Mittelwert von 5.17 signifikant positiver beurteilt werden als in Bezug auf ihre Verlässlichkeit und ihr Belastungspotential, die durchschnittlich mit einem Skalenwert von 4.54 bzw. 4.46 bewertet werden. Letztere erweisen sich nicht als substantiell unterschiedlich (Nützlichkeit vs. Verlässlichkeit: $t=6.32$, $df=138$; Nutzen vs. Belastung: $t=4.58$, $df=138$, beide: $p < 0.05$; Verlässlichkeit vs. Belastung: $t=0.50$, $df=138$; $p > 0.1$). Auffällig ist, dass sich der Konsens der Befragten über ihre Haltung zu Umfragen zwischen den Bewertungsdimensionen unterscheidet. So kann für alle Sponsoren festgestellt werden, dass die eingeschätzten Belastungen, gemessen an der Standardabweichung der Urteile, deutlich mehr Varianz aufweisen, als dies bei der wahrgenommenen Nützlichkeit und Zuverlässigkeit von Umfragen der Fall ist.

Insgesamt weisen die Teilnehmer unserer Studie ein hohes Ausmaß an Erfahrung mit Umfragen auf: 66 Prozent nahmen bereits mindestens einmal zuvor an einer Befragung teil. Informanten mit Umfragerfahrung haben vor der aktuellen Befragung durchschnittlich bereits an 4.0 Interviews teilgenommen. Werden auch Probanden ohne Umfragerfahrung einbezogen, so liegt dieser Wert bei 2.6 Befragungen.

Bei der Erfassung des Haushaltseinkommens haben bei der offenen Abfrage zuerst 20.1 Prozent der Probanden eine Antwort hierzu abgelehnt. Dieser Anteil konnte durch die kategorisierte Listenabfrage auf 10.1 Prozent reduziert werden.

4.2 Determinanten der generalisierten Umfrageeinstellung

Im folgenden Abschnitt wird untersucht, welche Bedeutung die Interviewerfahrung für den Einfluss der sponsorenspezifischen Bewertungsdimensionen auf die generalisierte Umfrageeinstellung hat. Es wird geprüft, ob der Zusammenhang zwischen den durchschnittlichen Bewertungen der Befragten auf allen neun Bewertungsdimensionen und der Umfrageeinstellung durch das Ausmaß der Umfragerfahrung moderiert wird.

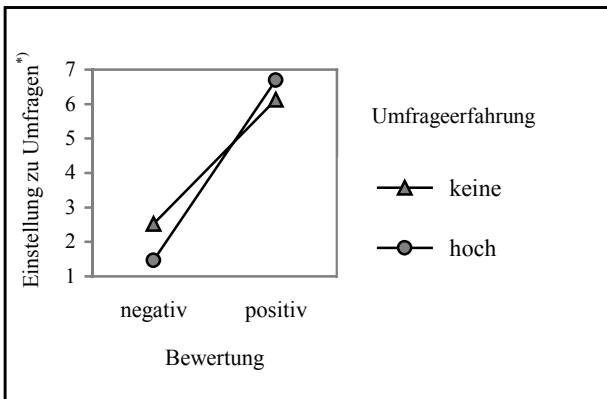
Tabelle 3: Bedeutung der Umfragerfahrung für die Erklärungskraft der aggregierten sponsorenspezifischen Bewertungsdimensionen (OLS Regressionsergebnisse)

	Modell 1 B (t-Wert)	Modell 2 B (t-Wert)
<i>Kontrollvariablen</i>		
Geschlecht (weiblich) ^{a)}	0.11 (1.83)	0.11 (1.24)
Fester Lebenspartner (ja) ^{b)}	0.41 (3.81)*	0.40 (3.79)*
Kirchgangshäufigkeit (hoch) ^{c)}	0.20 (1.83)	0.22 (2.03)*
Konfession (katholisch) ^{d)}	-0.21 (1.99)*	-0.21 (2.00)*
Länge der Schulbildung (Jahre)	-0.05 (1.72)	-0.05 (1.62)
<i>Bewertung / Umfragerfahrung</i>		
Durchschnittliche Bewertung (Skalenwerte)	0.69 (11.55)*	0.60 (8.09)*
Umfragerfahrung (Anzahl Umfragen)	0.00 (0.00)	-0.18 (2.03)*
Bewertung • Umfragerfahrung	--	0.04 (2.05)*
Konstante	1.50 (3.63)*	1.89 (4.20)*
Korrigiertes R ²	0.54	0.57
N	139	139

Signifikanz: * $p \leq 0.05$; Referenzkategorien: a) Männlich; b) Nein; c) Niedrig; d) Andere Konfession.

Die in Tabelle 3 dargestellten Ergebnisse zeigen, dass sich die durchschnittliche Bewertung der Umfragen unterschiedlicher Sponsoren in starkem Ausmaß auf die Umfrageeinstellung der Befragten auswirkt (vgl. Modell 1). Die Häufigkeit der Teilnahme an Umfragen in der Vergangenheit hat dagegen keinen Einfluss auf die Bewertung von Umfragen generell. Als wichtiges Ergebnis der vorliegenden Analyse zeigt sich in Modell 2, dass dem Ausmaß der Umfrageerfahrung eine bedeutsame Moderatorfunktion für den Zusammenhang zwischen den sponsorenspezifischen Bewertungen und der Umfrageeinstellung zukommt: Die Interaktion zwischen beiden Konstrukten ist eine statistisch abgesicherte Determinante der Umfrageeinstellung.

Abbildung 1: Bedeutung der Umfrageerfahrung für den Einfluss der durchschnittlichen sponsorenspezifischen Umfragebewertung auf die Umfrageeinstellung⁶



^{*)} Antwortskala von 1 (negative Einstellung) bis 7 (positive Einstellung)

In Abbildung 1 wird der Inhalt des festgestellten Interaktionseffektes aus dem Regressionsmodell 2 dargestellt. Es ist feststellbar, dass die Einstellung zu Umfragen von Probanden ohne vorherige Interviewerfahrung deutlich schwächer durch ihre Bewertungen der

6 Die Vorhersagewerte in Abbildung 1 beruhen auf dem in Tabelle 3 dargestellten Regressionsmodell 2. Sie gelten für Befragte ohne Umfrageerfahrung und solche mit einer Umfrageerfahrung von einer Standardabweichung über dem Durchschnitt der Stichprobe. Für die sponsorenspezifische Bewertung wurden die Werte 1 (negative Bewertung) und 7 (positive Bewertung) in die Prognosegleichung eingesetzt. Die Parameter der Kontrollvariablen sind auf der Referenzkategorie bzw. dem Populationsmittelwert fixiert.

Sponsoren beeinflusst wird, als dies bei Befragten mit einem hohen Ausmaß an Umfrageerfahrung in der Vergangenheit der Fall ist. Dieses Ergebnis kann als Hinweis dafür aufgefasst werden, dass die direkte Erfahrung mit dem Einstellungsgegenstand zu einer Kristallisation der Urteile und daher zu einer konsistenteren Bewertungsstruktur beim vorliegenden Thema führt.

In Tabelle 4 wird der Einfluss der nach den Sponsoren und Urteilkriterien differenzierten Bewertungen auf die Umfrageeinstellung, getrennt für Befragte mit und ohne Interviewerfahrung, dargestellt. Dabei zeigen sich deutliche Unterschiede darin, welche Einzelbewertungen sich in den unterschiedlich erfahrenen Befragtengruppen als relevant erweisen. So wirkt sich in der Gruppe der Befragten ohne Umfrageerfahrung ausschließlich die bewertete Nützlichkeit der Umfragen aller drei Sponsoren signifikant auf ihr Urteil über Umfragen insgesamt aus (vgl. Modell 1). Weder die bewertete Zuverlässigkeit noch die erwarteten Belastungen bei der Teilnahme an Umfragen der verschiedenen Sponsorentypen spielen hier eine Rolle.

Tabelle 4: Erklärungskraft der Bewertung von Umfragen unterschiedlicher Sponsoren hinsichtlich verschiedener Bewertungsdimensionen für die generalisierte Einstellung zu Umfragen (OLS-Regressionsergebnisse)

	Modell 1: Ohne Interviewerfahrung B (t-Wert)	Modell 2: Mit Interviewerfahrung B (t-Wert)	Modell 3: Alle B (t-Wert)
<i>Kontrollvariablen</i>			
Geschlecht (weiblich) ^{a)}	0.25 (1.39)	0.09 (0.72)	0.12 (1.31)
Fester Lebenspartner (ja) ^{b)}	0.33 (1.84)	0.43 (2.90)*	0.40 (3.61)*
Kirchgangshäufigkeit (hoch) ^{c)}	0.16 (0.84)	0.22 (1.47)	0.18 (1.58)
Konfession (katholisch) ^{d)}	-0.13 (0.60)	-0.25 (1.79)	-0.17 (1.51)
Schulbildung (Jahre)	0.02 (0.40)	-0.09 (2.36)*	-0.05 (1.58)
<i>Sponsor / Urteilsdimension</i>			
Kommerziell / Nützlichkeit	0.15 (2.63)*	0.08 (1.69)	0.11 (3.13)*
Wissenschaft / Nützlichkeit	0.17 (3.33)*	0.13 (2.44)*	0.13 (3.93)*
Staatlich / Nützlichkeit	0.10 (2.02)*	0.08 (1.71)	0.09 (2.75)*
Kommerziell / Verlässlichkeit	0.03 (0.56)	0.00 (0.08)	0.03 (0.87)
Wissenschaft / Verlässlichkeit	-0.08 (1.11)	0.15 (2.89)*	0.08 (2.10)*
Staatlich / Verlässlichkeit	0.09 (1.50)	0.04 (0.73)	0.04 (1.11)
Kommerziell / Belastung	0.04 (0.75)	0.08 (2.13)*	0.07 (2.36)*
Wissenschaft / Belastung	0.02 (0.31)	0.11 (2.78)*	0.08 (2.58)*
Staatlich / Belastung	0.03 (0.57)	0.04 (0.94)	0.05 (1.49)
Konstante	1.31 (1.84)	1.75 (3.08)*	1.44 (3.45)*
Korrigiertes R ²	0.62	0.52	0.55
N	47	92	139

Signifikanz: * $p \leq 0.05$; Referenz-Kategorien: ^{a)} Männlich; ^{b)} Nein; ^{c)} Niedrig; ^{d)} Andere;

Im Gegensatz dazu ziehen Befragte, die in der Vergangenheit bereits mehr oder weniger häufig an Befragungen teilgenommen haben, insbesondere ihre Bewertung von wissenschaftlichen Befragungen als Grundlage ihrer Einstellung zu Umfragen insgesamt heran: Die Bewertung dieses Sponsorentyps bezüglich aller drei Kriterien ist ein statistisch abgesicherter Prädiktor der Umfrageeinstellung (vgl. Modell 2). Abgesehen von den eingeschätzten Belastungen durch kommerzielle Umfragen erweist sich in dieser Befragtengruppe keine andere Bewertungsdimension als bedeutsame Determinante der Umfrageeinstellung. Bei der Analyse der Gesamtstichprobe erweist sich dann genau die Vereinigungsmenge der in den Teilgruppen relevanten Faktoren als statistisch signifikante Bestimmungsfaktoren der Einstellung zu Umfragen (vgl. Modell 3).

Die nach der Befragungserfahrung differenzierte Analyse der Bestimmungsfaktoren der Umfrageeinstellung erbringt ein weiteres interessantes Ergebnis. So ist erkennbar, dass die untersuchten Determinanten der Umfrageeinstellung bei Befragten ohne Umfrageerfahrung zu einem substanziell höheren korrigierten R^2 von 0.62 führen als dies mit einem R^2 -Wert von 0.52 bei Befragten mit Umfrageerfahrung der Fall ist. Dieser Unterschied in den Determinationskoeffizienten kann darauf zurückgeführt werden, dass durch die direkte Erfahrung mit dem Einstellungsgegenstand „Umfragen“ die Befragten über mehr für ihre Umfrageeinstellung relevante Informationen verfügen, die nicht durch die verwendeten Erklärungsfaktoren erfasst werden. Dieser Umstand führt bei bestehender Interviewerfahrung zu einem höheren Anteil unerklärter Varianz der generalisierten Umfrageeinstellung.

In Tabelle 5 wird dargestellt, wie sich der Anteil der erklärten Varianz der Umfrageeinstellung reduzieren würde, wenn man die Einzelbewertungen (für alle Kombinationen der 3 Sponsoren mit den 3 Urteilsdimensionen) bzw. alle Items zu einem Sponsor oder einer Urteilsdimension aus dem Erklärungsmodell entfernen würde. Dabei zeigt sich, dass einerseits der bewerteten Nützlichkeit von Befragungen und andererseits den Urteilen über Umfragen wissenschaftlicher Sponsoren die höchste Erklärungskraft zukommt: Die Reduktion der erklärten Varianz beträgt im ersten Fall 10.8 und im zweiten 8.8 Prozentpunkte. Nach den in Tabelle 4 dargestellten Ergebnissen geht die hohe Bedeutung der Nützlichkeitsbewertung auf Befragte ohne Umfrageerfahrung und die Erklärungskraft der Bewertung wissenschaftlicher Sponsoren auf Befragte mit Interviewerfahrung zurück. Bei der mit Abstand wichtigsten Einzelbewertung (R^2 -Reduktion=0.051), der Nutzenbewertung von Umfragen durch wissenschaftliche Sponsoren, handelt es sich um jenen Faktor, der sich sowohl für Befragte mit Umfrageerfahrung wie auch für solche ohne Befragungserfahrung als bedeutsamer Bestimmungsfaktor der Umfrageerfahrung erwiesen hat.

Tabelle 5: Einfluss der Bewertung von Umfragen unterschiedlicher Sponsoren nach verschiedenen Kriterien auf die generalisierte Umfrageeinstellung (alle Befragte)

	Kommerziell	Wissenschaftlich	Staatlich	Gesamt
Nützlichkeit	.032 *	.051 *	.025 *	.108
Verlässlichkeit	.002	.015 *	.004	.021
Belastung	.018 *	.022 *	.007	.047
Gesamt	.052	.088	.036	.176

Die Werte bezeichnen den Anstieg in R^2 , wenn bei Konstanthaltung aller anderen Faktoren die jeweilige Determinante in das Modell einbezogen wird. Signifikanz: * $p \leq 0.05$

4.3 Konsequenzen der Umfrageeinstellung für die Antwortbereitschaft

Nach der Analyse der Determinanten der Umfrageeinstellung soll nun gezeigt werden, ob diese auch Konsequenzen für das Verhalten der Befragten im Interview hat. Nach den Vorhersagen des MODE-Modells und des Modells der Frame-Selektion sollte die Erklärungskraft der Umfrageeinstellung für die Antwortverweigerung bei der Einkommensfrage mit der kognitiven Zugänglichkeit dieser Einstellungen zunehmen. Demnach sollte sich die Umfrageeinstellung in Interaktion mit den Antwortlatenzen als signifikanter Prädiktor der Wahrscheinlichkeit einer Antwortverweigerung erweisen. Diese Hypothese wird mit den in Tabelle 6 dargestellten Ergebnissen logistischer Regressionsanalysen überprüft.

Die Ergebnisse von Modell 1 zeigen, dass sich die Umfrageeinstellung alleine nicht als signifikanter Prädiktor der Verweigerung der Einkommensfrage erweist. In Modell 2 wird nun auch die durchschnittliche Antwortgeschwindigkeit bei der Beantwortung der Fragen über die Umfrageeinstellung berücksichtigt. Es zeigt sich, dass die Interaktion zwischen diesem Faktor und den inhaltlichen Angaben über die Einstellung zu Umfragen einen signifikanten Erklärungsbeitrag für die Vorhersage der Beantwortung oder Verweigerung der Einkommensfrage leistet.

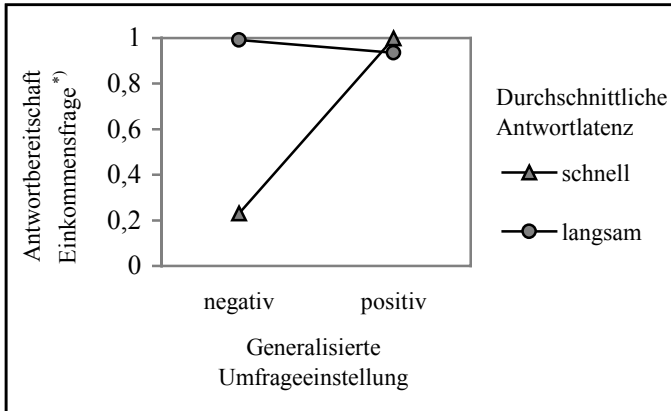
Tabelle 6: Erklärungskraft der Umfrageeinstellung für die Bereitschaft zur Beantwortung der Einkommensfrage (Ergebnisse logistischer Regressionsanalysen)

	Modell 1 B (Wald-Statistik)	Modell 2 B (Wald-Statistik)
<i>Kontrollvariablen</i>		
Geschlecht (weiblich) ^{a)}	-0.40 (0.44)	-0.37 (0.36)
Fester Lebenspartner (ja) ^{b)}	0.92 (1.13)	1.46 (2.10)
Kirchgangshäufigkeit (hoch) ^{c)}	0.85 (1.66)	-0.84 (1.34)
Konfession (katholisch) ^{d)}	-0.77 (0.86)	-1.14 (1.42)
Länge der Schulbildung (Jahre)	-0.20 (1.20)	-0.23 (1.37)
<i>Umfrageeinstellung / Antwortlatenz</i>		
Umfrageeinstellung (Skalenwerte)	0.45 (1.14)	1.62 (5.64)*
Antwortlatenz (Rohskalenwerte)	--	1.10 (4.98)*
Umfrageeinstellung • Antwortlatenz	--	-0.24 (5.52)*
Konstante	3.15 (1.14)	-1.52 (0.16)
McFaddens Pseudo R ²	0.10	0.16
N	139	139

Signifikanz: * $p \leq 0.05$; Referenzkategorien: ^{a)} Männlich; ^{b)} Nein; ^{c)} Niedrig; ^{d)} Andere Konfession; AV: 0=Verweigerung der Einkommensfrage, 1=Beantwortung der Einkommensfrage (offen oder anhand der vorgegebenen Kategorien) sowie „weiß nicht“.

Der Inhalt der Interaktion zwischen der Umfrageeinstellung und deren Antwortlatenz lässt sich durch die auf der Grundlage von Regressionsmodell 2 prognostizierten Wahrscheinlichkeit der Beantwortung der Einkommensfrage veranschaulichen. Diese Prognosen sind in Abbildung 2 für unterschiedliche Ausprägungen der Umfrageeinstellung und der Geschwindigkeit bei deren Beantwortung dargestellt. Dabei lässt sich feststellen, dass sich die inhaltlichen Angaben über die Umfrageeinstellung nur für Befragte mit schnellen Antwortlatenzen und damit einer starken Einstellungszugänglichkeit auf die Wahrscheinlichkeit einer Antwortverweigerung auswirken. Für diese Gruppe mit überdurchschnittlich schnellen Antwortlatenzen wird bei einer sehr positiven Umfrageeinstellung beinahe mit Sicherheit die Beantwortung der Einkommensfrage prognostiziert, wohingegen diese Wahrscheinlichkeit bei sehr negativer Einstellung auf 23 Prozent zurückgeht. Dagegen erweisen sich die Einstellungsangaben von Probanden mit geringer Einstellungsstärke als weitgehend irrelevant für die Wahrscheinlichkeit einer Antwortverweigerung: Die Wahrscheinlichkeit einer Antwort liegt bei positiver Umfrageeinstellung bei 93 Prozent und bei negativer Einstellung bei 99 Prozent.

Abbildung 2: Erklärungskraft der generalisierten Umfrageeinstellung und deren Antwortlatenzen für die Beantwortung der Einkommensfrage ⁷



^{*)} Wahrscheinlichkeit der Nennung des Haushaltseinkommens

Interessanterweise lässt sich demnach für Befragte mit einer kognitiv wenig verankerten Umfrageeinstellung, unabhängig von deren Inhalt, eine generell hohe Bereitschaft zur Beantwortung der Einkommensfrage feststellen. Mit zunehmender Einstellungsstärke steigt dann die Handlungsrelevanz einer negativen Haltung zu Umfragen: Nur für Befragte mit einer negativen und gleichzeitig kognitiv stark verankerten Umfrageeinstellung findet sich ein starkes Kooperationsdefizit. Dieser Effekt lässt sich so interpretieren, dass das Verhalten von Befragten ohne wirksame Umfrageeinstellung durch eine generalisierte kooperative Grundhaltung geprägt wird. Die rahmende Wirkung dieser, für das Verhalten in Umfragen unspezifischen Einstellung, wird erst dann durch die der Einstellung zu Umfragen abgelöst, wenn deren kognitive Zugänglichkeit zunimmt.

7 Die Vorhersagewerte in Abbildung 2 gelten für Befragte mit einer Standardabweichung über und unter dem Mittelwert der durchschnittlich beobachteten Antwortlatenzen. Für die Umfrageeinstellung wurden die Werte 1 (negative Einstellung) und 7 (positive Einstellung) in die Prognosegleichung eingesetzt. Die Parameter der Kontrollvariablen sind auf der Referenzkategorie oder dem Populationsmittelwert fixiert.

5. Zusammenfassung und Diskussion

In dieser Studie wurden zuerst explorativ die Determinanten der generalisierten Einstellung zu Umfragen untersucht. Im zweiten Schritt wurden dann, auf der Grundlage von Prognosen des MODE-Modells und des Modells der Frame-Selektion, Hypothesen darüber getestet, unter welchen Umständen die Umfrageeinstellung Konsequenzen für die Bereitschaft der Befragten zur Beantwortung der Einkommensfrage hat.

Nach den vorliegenden Ergebnissen haben die Teilnehmer unserer Studie eine insgesamt positive generalisierte Einstellung zu Umfragen, wobei sich aber auch eine deutliche Differenzierung in der Bewertung von Befragungen hinsichtlich verschiedener Sponsorentypen und Urteilkriterien feststellen lässt: Wissenschaftliche Befragungen und die Nützlichkeit von Umfragen werden am stärksten anerkannt. Unsere Analyse hat auch gezeigt, dass sich diese beiden Bewertungsdimensionen am stärksten auf die Einstellung der Befragten zu Umfragen generell auswirken. Bei einer nach dem Ausmaß der Umfrageerfahrung der Probanden differenzierte Analyse dieser Zusammenhänge hat sich gezeigt, dass Befragte ohne Interviewerfahrung ausschließlich die bewertete Nützlichkeit und solche mit Umfrageerfahrung primär ihre Bewertung wissenschaftlicher Befragungen zur Grundlage ihrer generalisierten Umfrageeinstellung machen. Ein weiteres Ergebnis ist, dass sich bei Befragten mit Umfrageerfahrung, verglichen mit Probanden ohne Interviewerfahrung, die sponsorenspezifischen Bewertungen signifikant stärker auf die Umfrageeinstellung auswirken. Mit dem Anteil der dadurch bei der Umfrageeinstellung erklärten Varianz verhält es sich jedoch genau umgekehrt. Demnach führt die direkte Erfahrung mit dem Einstellungsgegenstand „Umfragen“ einerseits zu einer stärkeren Strukturierung des gesamten Einstellungsfeldes, zugleich aber auch zu einem höheren Anteil an Einstellungsvariabilität, der nicht durch die sponsorenspezifischen Bewertungen erfasst wird.

Es hat sich gezeigt, dass die generelle Umfrageeinstellung einen klaren Einfluss auf die Wahrscheinlichkeit kooperativen Verhaltens der Befragten während des Interviews ausübt. So unterscheidet sich die Wahrscheinlichkeit für die Beantwortung der Frage nach dem Netto-Haushaltseinkommen deutlich zwischen Probanden mit einer positiven und negativen Einstellung zu Umfragen. In Übereinstimmung mit den theoretischen Vorhersagen lässt sich dieser Effekt aber nur dann feststellen, wenn eine relativ schnelle Beantwortung der Fragen über die generalisierte Umfrageeinstellung eine hohe kognitive Zugänglichkeit dieser Einstellung anzeigt. Demnach werden die „latenten“ Urteile der Befragten über Umfragen zunehmend für die Kooperationsbereitschaft und das konkrete Verhalten im Interview relevant, wenn die Einstellungsstärke ansteigt. Somit kommt dem Inhalt und der

Stärke der Umfrageeinstellung sowie deren Zusammenspiel eine große Bedeutung für die Wahrscheinlichkeit von Item-Nonresponse und damit der Qualität von Umfragedaten zu.

In früheren Arbeiten hat sich die Qualität von Erfahrungen der Befragten mit Interviews in der Vergangenheit als eine weitere wichtige Determinante der generalisierten Einstellung zu Umfragen erwiesen (Gove/Geerken 1977; Stocké/Langfeldt 2004). Demnach bewirken negative Interviewerfahrungen eine kritische Haltung der Befragten zu Umfragen und nach unseren Ergebnissen dann eine erhöhte Wahrscheinlichkeit für die Verweigerung von Angaben in späteren Befragungen. Umfrageforscher sollten daher bei der Gestaltung der Interviews, etwa in Hinblick auf deren Länge und den Schwierigkeitsgrad der Fragen, im Auge behalten, dass sich der Grad der Belastung der Befragten auf deren Umfrageeinstellung und damit auf die zukünftig realisierbare Datenqualität auswirkt.

Nach unseren Ergebnissen ist die Bewertung von Befragungen durch unterschiedliche Typen von Sponsoren eine bedeutsame Determinante der generalisierten Umfrageeinstellung. Diese hat sich in der vorliegenden Untersuchung auf die Antwortbereitschaft eines wissenschaftlichen Sponsors ausgewirkt. Aus beiden Ergebnissen zusammengenommen kann gefolgert werden, dass die Bewertung bestimmter Sponsorentypen, vermittelt über den Einfluss auf die generalisierte Umfrageeinstellung, die Kooperationsbereitschaft bei Befragungen durch ganz andere Sponsoren beeinflusst. Somit tragen Umfrageforscher nicht nur Verantwortung für die Möglichkeit einer erfolgreichen Feldarbeit der Organisationen des eigenen Sponsorentyps, sondern beeinflussen auch die erreichbare Datenqualität bei Umfragen ganz anderer Sponsoren.

Die vorliegende Studie wurde von einer wissenschaftlichen Organisation und damit von einem Sponsorentyp durchgeführt, dessen Bewertung sich als eine der wichtigsten Determinanten der generalisierten Umfrageeinstellung herausgestellt hat. Dabei muss die Frage gestellt werden, ob diese Übereinstimmung möglicherweise darauf zurückgeht, dass die spezielle Sponsorenschaft der Untersuchung die entsprechenden Bewertungen situational aktiviert und somit kognitiv verfügbar gemacht hat. Unser Ergebnis, wonach nur die Umfrageeinstellung von Befragten mit Umfrageerfahrung überhaupt durch die Bewertung wissenschaftlicher Sponsoren beeinflusst wird, macht eine generelle Erklärung der relativen Erklärungskraft der Sponsorenbewertungen durch Unterschiede im situationalen Aktivierungsniveau unwahrscheinlich. Eine Replikation unserer Studie durch einen alternativen Sponsor wäre für die Absicherung der Ergebnisse dennoch wünschenswert.

Da die vorliegende Datengrundlage nur auf einer Zufallsstichprobe einer Großstadt beruht, lassen sich die gefundenen Ergebnisse über die Randverteilungen der Umfrageeinstellung und der spezifischen Bewertungen von Umfragen nicht einfach auf bundesdeutsche Verhältnisse übertragen. Es ist bekannt, dass in Großstädten überproportional viele

Marktforschungsinstitute und Forschungseinrichtungen aufzufinden sind, was zu einem besonders hohen Ausmaß an Interviewerfahrung führt. Es ist denkbar, dass die Ergebnisse in Bevölkerungsteilen mit weniger intensiver Befragungstätigkeit anders ausgefallen wären. Daher sollte unsere Studie mit den Daten einer bevölkerungsrepräsentativen Stichprobe und einer größeren Fallzahl repliziert werden.

Problematisch ist weiterhin die Selektivität der Umfrageteilnehmer nach dem Untersuchungsgegenstand. Da die Ergebnisse anderer Studien darauf hindeuten, dass die größten Umfrageskeptiker erst gar nicht an Befragungen teilnehmen, lassen sich die gefundenen Ergebnisse nicht ohne Weiteres auf die Gruppe der Teilnahmeverweigerer generalisieren. Diese Einschränkung trifft allerdings nur für die im ersten Teil unserer Untersuchung gefundenen Resultate über die Determinanten der generalisierten Umfrageeinstellung zu. Da sich die Schlussfolgerungen über die Bedeutung der Umfrageeinstellung für die Kooperationsbereitschaft im Interview naturgemäß nur auf Befragungsteilnehmer beziehen, bleiben diese hiervon unberührt.

Literatur

Albaum, G., 1987: Do Source and Anonymity Affect Mail Survey Results? *Academy of Marketing Science* 15: 74-81.

Bassili, J.N., 1993: Response Latency versus Certainty as Indexes of the Strength of Voting Intentions in a CATI Survey. *Public Opinion Quarterly* 57: 54-61.

Bassili, J.N., 1995: Response Latency and the Accessibility of Voting Intentions: What Contributes to Accessibility and How it Affects Vote Choice. *Personality and Social Psychology Bulletin* 21: 686-695.

Bassili, J.N./Bors, D.A., 1997: Using Response-Latency to Increase Lead Time in Election Forecasting. *Canadian Journal of Behavioural Science* 29: 231-238.

Doob, A.N./Freedman, J.L./Carlsmith, J.M., 1973: Effects of Sponsor and Prepayment on Compliance with a Mailed Request. *Journal of Applied Psychology* 57: 346-347.

Dran, E.M./Hildreth, A., 1995: What the Public Thinks About How we Know What it is Thinking. *International Journal of Public Opinion Research* 7: 128-144.

Erbslöh, B./Koch, A., 1988: Die Non-Response-Studie zum Allbus 1986: Problemstellung, Design, erste Ergebnisse. *ZUMA-Nachrichten* 22: 29-44.

Esser, H., 2001: *Soziologie. Spezielle Grundlagen. Band 6: Sinn und Kultur*. Frankfurt a. M.: Campus.

Fazio, R.H., 1990: Multiple Processes by which Attitudes Guide Behavior: The Mode Model as an Integrative Framework. S.75-104 in: Zanna, M.P. (Ed.), *Advances in Experimental Social Psychology*. San Diego: Academic Press.

- Fazio, R.H./Powell, M.C./Williams, C.J., 1989: The Role of Attitude Accessibility in the Attitude-to-Behavior Process. *Journal of Consumer Research* 16: 280-288.
- Fazio, R.H./Williams, C.J., 1986: Attitude Accessibility as a Moderator of the Attitude-Perception and Attitude-Behavior Relations: An Investigation of the 1984 Presidential Election. *Journal of Personality and Social Psychology* 51: 505-514.
- Fletcher, J.F., 2000: Two Timing: Politics and Response Latencies in a Bilingual Survey. *Political Psychology* 21: 27-55.
- Forsa, 1991: Umfragen in Deutschland. Erfahrungen und Meinungen der Bürger 1991. *Context* 8: 1-14.
- Forsa, 1993a: Umfragen in Deutschland. Erfahrungen und Meinungen der Bürger. Teil I. *Context* 7: 1-7.
- Forsa, 1993b: Umfragen in Deutschland. Erfahrungen und Meinungen der Bürger. Teil II. *Context* 8: 1-10.
- Forsa, 1996: Umfragen über Umfragen: Ein Jahrzehnt im Überblick. *Context* 14: 1-15.
- Forsa, 2000: Die Akzeptanz von Umfragen in Deutschland Anfang 2000. *Context* 7: 2-13.
- Gove, W.R./Geerken, M.R., 1977: Response Bias in Surveys of Mental Health: An Empirical Investigation. *American Journal of Sociology* 82: 1289-1317.
- Goyder, J., 1986: Surveys on Surveys: Limitations and Potentialities. *Public Opinion Quarterly* 50: 27-41.
- Hawkins, D.I., 1979: The Impact of Sponsor Identification and Direct Disclosure of Respondent Rights on the Quantity and Quality of Mail Survey Data. *Journal of Business* 52: 577-590.
- Houston, M.J./Nevin, J.R., 1977: The Effects of Source and Appeal on Mail Survey Response Patterns. *Journal of Marketing Research* 14: 374-378.
- Jones, W.H./Lang, J.R., 1983: Reducing Item Omission in Mail Surveys. *Political Methodology* 9: 285-293.
- Kohut, A., 1986: Rating the Polls: The Views of Media Elites and the General Public. *Public Opinion Quarterly* 50: 1-9.
- Kokkinaki, F./Lunt, P., 1997: The relationship between involvement, attitude accessibility and attitude-behaviour consistency. *British Journal of Social Psychology* 36: 497-509.
- Leiblein, A./Klass, J., 1990: Die Akzeptanz von Umfragen. *Planung & Analyse* 7: 259-262.
- McDaniel, S.W./Verille, P./Madden, C.S., 1985: The Threats to Marketing Research: An Empirical Reappraisal. *Journal of Marketing Research* 21: 74-80.
- Orr, C.J., 1976: Sponsorship, Self Presentation, and Legitimacy. *Communication Monographs* 43: 80-90.

- Peterson, R.A., 1975: An Experimental Investigation of Mail-Survey Responses. *Journal of Business Research* 3: 199-210.
- Porst, R., 1998: Erfahrung mit und Bewertung von Umfragen. Was unsere Befragten über Umfragen denken. ZUMA-Arbeitsbericht 3: 1-14.
- Presser, S./Blair, J./Triplett, T., 1992: Survey Sponsorship, Response Rates, and Response Effects. *Social Science Quarterly* 73: 699-702.
- Reuband, K.-H., 1991: Bekundete und reale Offenheit im Interview. *Planung & Analyse* 18: 49-51.
- Rogelberg, S.G./Fisher, G.G./Maynard, D.C./Hakel, M.D./Horvath, M., 2001: Attitudes Toward Surveys: Development of a Measure and its Relationship to Respondent Behavior. *Organizational Research Methods* 4: 3-25.
- Rogelberg, S.G./Conway, J.M./Sederburg, M.E./Spitzmüller, C./Aziz, S./Knight, W.E., 2003: Profiling Active and Passive Nonrespondents to an Organizational Survey. *Journal of Applied Psychology* 88: 1104-1114.
- Schöne, D., 1973: Auftragbereifekt und Erfolg von Mahnschreiben bei einer postalischen Befragung. in: Erbslöh, E./Esser, H./Reschka, W./Schöne, D. (Hrsg.), *Studien zum Interview*. Meisenheim am Glan: Verlag Anton Hain.
- Schützenmeister, F., 2002: Die Bereitschaft, sich wieder befragen zu lassen, in postalischen Erhebungen. *Zeitschrift für Soziologie* 31: 138-154.
- Schleifer, S., 1986: Trends in Attitudes Toward and Participation in Survey Research. *Public Opinion Quarterly* 50: 17-26.
- Sharp, L.M./Frankel, J., 1983: Respondent Burden: A Test of Some Common Assumptions. *Public Opinion Quarterly* 47: 36-53.
- Six, B./Eckes, T., 1996: Metaanalysen in der Einstellung-Verhaltens-Forschung. *Zeitschrift für Sozialpsychologie* 27: 7-17.
- Stadt Mannheim, 1998: Elektronischer Datensatz über die Verteilung der Bevölkerungsmerkmale im Stadtgebiet, Stand 31.12.1998. Mannheim: Statistikstelle.
- Stinchcombe, A.L./Jones, C./Sheatsley, P., 1981: Nonresponse Bias for Attitude Questions. *Public Opinion Quarterly* 45: 359-375.
- Stocké, V., 2004 : Entstehungsbedingungen von Antwortverzerrungen durch soziale Erwünschtheit. Ein Vergleich der Prognosen der Rational-Choice Theorie und des Modells der Frame-Selektion. *Zeitschrift für Soziologie*, im Druck.
- Stocké, V./Langfeldt, B., 2004: Effects of Survey Experience on Respondents' Attitudes towards Surveys. *Bulletin de Methodologie Sociologique* 81: 5-32.

Anhang

Instrument zur Erfassung der Bewertung von Umfragen unterschiedlicher Sponsoren bezüglich verschiedener Kriterien

In Deutschland und besonders in Mannheim werden viele Befragungen über Politik, Konsumprodukte und sozialwissenschaftliche Themen durchgeführt. Über den Sinn von solchen Umfragen kann man ganz unterschiedlicher Ansicht sein. Ich lese Ihnen nun einige Meinungen zu diesem Thema vor. Würden Sie mir bitte in jedem Fall sagen, in welchem Ausmaß Sie ganz persönlich den Aussagen zustimmen können. [Antwortskala wird vorgelegt]. Sie können Ihre Meinung auf dieser Skala zwischen dem Wert 1 „lehne voll und ganz ab“ und dem Wert 7 „stimme voll und ganz zu“ abstufen. Nennen Sie bitte immer die Zahl, die Ihrer Zustimmung entspricht.

1. Umfragen von kommerziellen Meinungsforschungsinstituten sind nützlich
2. Umfragen von wissenschaftlichen Forschungseinrichtungen sind nützlich
3. Umfragen von statistischen Ämtern und staatlichen Behörden sind nützlich
4. Kommerzielle Umfrageergebnisse sind zuverlässig
5. Wissenschaftliche Umfrageergebnisse sind zuverlässig
6. Staatliche Umfrageergebnisse sind zuverlässig
7. Die Teilnahme an kommerziellen Umfragen ist unangenehm
8. Die Teilnahme an wissenschaftlichen Umfragen ist unangenehm
9. Die Teilnahme an staatlichen Umfragen ist unangenehm

Korrespondenzadressen

Dr. Volker Stocké

Sonderforschungsbereich 504

„Rationalitätskonzepte, Entscheidungsverhalten und ökonomische Modellierung“

TP 7 „Bildungsaspirationen, Bezugsgruppen und Bildungsentscheidungen“

Universität Mannheim

L13, 15

D-68131 Mannheim

Tel.: 0621-1813432

vstocke@rumms.uni-mannheim.de

Dipl. Soz. Birgit Becker

Lehrstuhl für Soziologie und Wissenschaftslehre

Universität Mannheim

Seminargebäude A 5

D-68131 Mannheim

Tel.: 0621-1812024

bbecker@rumms.uni-mannheim.de

COMPUTERUNTERSTÜTZTE INHALTSANALYSE OHNE DIKTIONÄR?

Ein Praxistest

JULIANE LANDMANN & CORNELIA ZÜLL

Der Artikel geht der Frage nach, ob computerunterstützte Inhaltsanalysen, die auf Co-Occurrence-Ansätzen basieren, eine Alternative zu diktionärbasierten Ansätzen darstellen. Nach einer kurzen Beschreibung der Vorgehensweise beim Arbeiten mit den beiden Ansätzen werden die häufig genannten Vorteile des Co-Occurrence-Ansatzes mit einem Praxistest überprüft. Anhand der Frage nach Unterschieden in der amerikanischen und deutschen Berichterstattung über den Irakkrieg wird mit Hilfe des Softwareprogramms Catpac die für Co-Occurrence-Ansätze typische Vorgehensweise Schritt für Schritt praktiziert und entsprechend detailliert beschrieben. Zum Schluss werden die aufgetretenen Vor- und Nachteile des Ansatzes zusammengefasst und diskutiert.

The paper contributes to the discussion whether computer-assisted content analysis based on co-occurrence analysis can replace dictionary based-approaches. Both approaches are described briefly before we turn to consider frequently cited advantages of the co-occurrence approach. Coverage of the war in Iraq in German and American newspapers provides the basis for a demonstration of the co-occurrence approach. Using the program Catpac, a typical analysis procedure is described step by step. Finally, the advantages and disadvantages of the approach are summarized and discussed.

1. Einleitung

Die Verwendung der computerunterstützten Inhaltsanalyse (cui) als wissenschaftliche Methode hat in den Sozialwissenschaften eine lange Tradition. Bereits in den 1960-er Jahren wurden die ersten Computerprogramme entwickelt¹ und 1967 fand die erste Kon-

1 Zum Beispiel die Programme General Inquirer (Stone et al. 1966) und WORDS (Iker/Harway 1969).

ferenz zum damaligen Forschungsstand der *cui* statt². Parallel zum Einzug des alltäglichen Gebrauchs von Computern ist erneut Bewegung in die Diskussion um die Anwendung computerunterstützter Inhaltsanalysen für sozialwissenschaftliche Fragestellungen gekommen. Diskussionsbedarf besteht vor allem zu Möglichkeiten, die verbreitete, aber sehr aufwendige Vorgehensweise des diktionärbasierten Ansatzes adäquat zu ersetzen oder zumindest zu vereinfachen.

Im vorliegenden Artikel wird diese Diskussion aufgegriffen und versucht, anhand eines Praxistests eine der diskutierten Alternativen zu bewerten. Vor dem Praxistest wird zunächst eine kurze Einführung in die typische Vorgehensweise bei diktionärbasierten Ansätzen und bei der hier vorgestellten Alternative gegeben. Ferner werden die methodischen Kriterien für die angestrebte Bewertung des zweiten Ansatzes präzisiert.

2. Ansätze der computerunterstützten Inhaltsanalyse

Für die Verwendung von Computern zur Erfassung von Textinhalten in den Sozialwissenschaften existiert eine Vielzahl unterschiedlicher Vorgehensweisen beziehungsweise Ansätze (siehe Alexa 1997). Der am häufigsten verwendete Ansatz ist der diktionärbasierte Ansatz (Züll/Landmann 2002).

Kern des diktionärbasierten Vorgehens ist, wie der Name verrät, ein Diktionär (Wörterbuch). Anhand einer Hypothese werden in einem ersten Schritt Erwartungen bezüglich des Inhalts der zu untersuchenden Texte formuliert, die dann im Rahmen eines Diktionärs als Kategorien operationalisiert werden. Dieses Diktionär enthält Kategorien, die durch Wort- und Phrasenlisten definiert werden. Beispiele für Wörter einer Wortliste der Kategorie „Nationen“ sind Deutschland, Frankreich, Italien, Kanada und USA. Als eine hierzu passende Phrase ist beispielsweise „amerikanisches Volk“ denkbar.

Mit Hilfe der den jeweiligen Kategorien zugeordneten Wort- und Phrasenlisten, und in einigen Fällen auch Codierregeln, wird die Zuordnung der Codes zu Textelementen (Wörter und Phrasen) organisiert. Werden zusätzlich Codierregeln verwendet, wird zum Beispiel festgelegt, welche Wörter gemeinsam in einem bestimmten Abstand, zum Beispiel innerhalb von fünf Wörtern, vorkommen müssen und welche Wörter nicht gemeinsam auftreten dürfen.

Während Programme wie TEXTPACK (Mohler/Züll 2001) ausschließlich mit Wort- und Phrasenlisten arbeiten, werden zum Beispiel in Programmen wie General Inquirer und

2 Auf der Annenberg-School-Content-Analysis-Konferenz in Philadelphia wurden Konzepte und methodische Überlegungen zur *cui* ausgetauscht und diskutiert (Gerbner et al. 1969).

Keds (Schrodt/Davis/Wedde 1994) Codierregeln hinzugenommen. Die anhand von Kategorien generierten Codes werden in der Regel mit statistischen Verfahren weiter analysiert.

Eine gegenwärtig an Bedeutung gewinnende Alternative zu den dictionärbasierten Ansätzen sind die sogenannten Co-Occurrence-Ansätze. Sie benötigen keine Definition von Kategorien und damit auch kein Diktionär im klassischen Sinn. Hier wird das gemeinsame Auftreten von Wörtern untersucht und durch Verfahren wie zum Beispiel Clusteranalysen oder MDS werden Assoziationsmuster analysiert. Aus dem gemeinsamen Auftreten/Nichtauftreten von Wörtern innerhalb eines Textfensters wird eine Ähnlichkeitsmatrix errechnet, die die Basis für die spätere Klassifikation bildet.

3. Muss es immer ein Diktionär sein? Probleme und Alternativen

Zur Anwendung des dictionärbasierten Ansatzes müssen zwei Voraussetzungen erfüllt sein: In der Regel liegen aus einer Theorie deduzierte Hypothesen der Definition von Kategorien zugrunde und darauf aufbauend wird ein Diktionär entwickelt und validiert. Dies bedeutet einen enormen Aufwand, vor allem wenn man bedenkt, dass die meisten Diktionäre nur zum einmaligen Gebrauch entwickelt werden. Deshalb wird immer häufiger nach Ansätzen gefragt, die die Diktionärkonstruktion vereinfachen oder gar unnötig machen. Eine Möglichkeit ist die Verwendung von bereits bestehenden Diktionären. Die Verwendung solcher Diktionäre setzt jedoch einiges voraus:

1. Zunächst muss ein der Fragestellung angemessenes Diktionär zur Verfügung stehen, das alle erforderlichen Kategoriendefinitionen beinhaltet.
2. Zudem muss bei der Verwendung bestehender Diktionäre sichergestellt werden, dass sie aktuell sind, das heißt, regelmäßig gewartet werden.
3. Desweiteren muss gewährleistet werden, dass das Diktionär im ausgewählten Programm korrekt verwendet wird, denn es ist nicht nur entscheidend, welche Kategorien definiert sind, sondern auch, wie sie vom Programm interpretiert und Texten zugewiesen werden. Zum Beispiel kann in einem Programm die Codierung von Phrasen Vorrang vor der Codierung von einzelnen Wörtern haben, in einem anderen Programm jedoch nicht. Dies kann dann zu unterschiedlichen Codierungen und damit zu unterschiedlichen Ergebnissen führen.

Sind alle Voraussetzungen zur Verwendung eines fertigen Diktionärs erfüllt, ist der Einsatz eines bereits verfügbaren, validierten Diktionärs sicher der am wenigsten arbeitsintensive Weg. Ist dies aber nicht gegeben, versprechen Co-Occurrence-Ansätze,

die keine im Voraus definierten Kategorien benötigen, schnelle und einfache Analysemöglichkeiten. Als ein weiterer Vorteil dieser Art von Ansätzen wird die offene Herangehensweise an den Text betont. Im Gegensatz zu den diktionsbasierten Vorgehensweisen bestimmt bei Co-Occurrence-Ansätzen der vorliegende Text die Themen, die analysiert werden, und nicht eine vorher aus einer Theorie deduzierte Hypothese (siehe dazu Salisbury 2001:68 und Hogenraad/McKenzie/Péladeau 2003).

Für uns stellt sich die Frage, ob und in wieweit Co-Occurrence-Ansätze tatsächlich eine Alternative zu den diktionsbasierten Ansätzen sind. Im folgenden Praxistest liegt der Schwerpunkt des Interesses folglich auf der Untersuchung des Arbeitsaufwands und den methodischen Chancen der offenen Herangehensweise.

Es existiert eine breite Palette von Anwendungsvarianten unter dem Oberbegriff Co-Occurrence-Analyse. Diese Vielfalt spiegelt sich auch auf Programmebene wider. Abhängig von der Wahl des Programms kommen verschiedene Abstandsmaße und unterschiedliche Klassifikationsverfahren zur Anwendung. Zur Berechnung von Co-Occurrences wird entweder das gemeinsame Auftreten oder Nichtauftreten von Wörtern in einem Textfenster gemessen oder es werden neuronale Netzwerke zur Berechnung verwendet. Als Klassifikationsverfahren werden in der Regel Clusteranalysen oder multidimensionale Skalierungsverfahren eingesetzt. Programmeispiele sind Catpac (Woelfel 1998), Hamlet (Brier 2003) und TextSmart (SPSS Inc. 1997).

Für unsere Analysen wählten wir Catpac aus, da es sich bei Catpac um ein relativ häufig verwendetes Programm³ handelt. Catpac arbeitet mit neuronalen Netzen, um die Relationen zwischen den häufigsten Wörtern zu berechnen, die in der daran anschließenden hierarchischen Clusteranalyse und multidimensionalen Skalierung als Ähnlichkeiten verwendet werden. Im Gegensatz zu den einfachen Co-Occurrence-Ansätzen, die lediglich über das gemeinsame Auftreten von Wörtern in einem Textfenster Rückschlüsse bezüglich ihrer jeweiligen Fragestellung ziehen, ist der in Catpac angewendete Ansatz der neuronalen Netze eine komplexere und detailliertere Variante des Co-Occurrence-Ansatzes (Rosen et al. 2003:5).

Ziel des folgenden Praxistests ist, eine konkrete Basis für eine methodische Diskussion zu schaffen. Die Ergebnisse des hier vorgestellten Beispiels selbst sollten daher inhaltlich nicht überbewertet werden.

3 Anwendungsbeispiele finden sich in (Freeman/Barnett 1994), (Salisbury 2001) und (Sherblom/Reinsch/Beswick 2001).

4. Co-Occurrence-Analyse

4.1 Fragestellung und Texte

Für unser Anwendungsbeispiel wählten wir die folgende Fragestellung: Welche Themen standen im Mittelpunkt der deutschen und der amerikanischen Medienberichterstattung über den Irakkrieg und welche relationalen Unterschiede sind bei der Themenbearbeitung zu beobachten?

Die Textbasis zur Beantwortung dieser Frage rekrutiert sich aus zwei überregional publizierten Tageszeitungen, die typische Beispiele der amerikanischen und deutschen Presse sind: Als Beispiel für die amerikanische Berichterstattung fungiert die New York Times und für die deutsche Berichterstattung die Frankfurter Allgemeine Zeitung. Berücksichtigung fanden alle Artikel, die zu Anfang oder zum Ende des Irakkrieges in der Rubrik „Politik“ der Frankfurter Allgemeinen Zeitung beziehungsweise in den Rubriken „International“ und „National & Politics“ der New York Times veröffentlicht wurden und das Wort „Irak“ oder „Iraq“ beinhalteten. Alle verwendeten Artikel wurden über die Lexis-Nexis-Datenbank⁴ bezogen. Die für den Anfang des Krieges gesammelten Artikel stammen aus der Zeit zwischen dem 18. und 20. März 2003 und die für das Ende gesammelten stammen aus der Zeit zwischen dem 7. und 9. April 2003. In der Nacht vom 19. zum 20. März begannen die Bombardements auf den Irak und am 9. April wurde Bagdad durch die alliierte Armee eingenommen. Es war der 21. Tag der Intervention und mit der Einnahme Bagdads wurde das inoffizielle Ende des Krieges in den Medien verkündet. Insgesamt besteht unsere Textbasis aus 435 Artikeln. Die detaillierte Darstellung der Verteilung der Artikel auf die zwei Zeitungen und auf die zwei unterschiedlichen Zeitpunkte ist in Tabelle 1 zu finden.

Tabelle 1: Verteilung der Artikel pro Tageszeitung

	Kriegsbeginn 18.03.-20.03.2003		Kriegsende 07.04.-09.04.2003	
	FAZ	NYT	FAZ	NYT
Anzahl der Artikel	131	107	84	113
Anzahl der Wörter	ca. 68.000	ca. 84.000	ca. 38.800	ca. 97.000

4 Die Lexis-Nexis-Datenbank ist online zugänglich (<http://www.lexis-nexis.com/>) und bietet unter anderen eine Vielzahl von Tageszeitungen als Volltext an.

Ohne bereits Aussagen über den Inhalt der Medienberichterstattung machen zu können, fällt bei Betrachtung der Tabelle 1 auf, dass rein quantitativ die Ausführlichkeit der Berichterstattung zwischen beiden Zeitungen differiert. Entsprechend verweist die Anzahl der Wörter in Relation zu der Anzahl der Artikel auf eine ausführlichere Berichterstattung in der New York Times. Des Weiteren fällt auf, dass in der Frankfurter Allgemeinen Zeitung zum Kriegsende hin das Ausmaß der Berichterstattung über den Krieg deutlich weniger umfangreich war als zu Kriegsbeginn, während in der New York Times der Umfang der Berichterstattung zugenommen hatte.

4.2 Textvorbereitung

Nachdem die Textbasis zusammengestellt war, wurden Vorarbeiten für die Inhaltsanalyse mit Catpac durchgeführt:

1. Um zu markieren, wann eine neue Einheit beginnt, wurde als erstes, den Anforderungen von Catpac entsprechend, jedes Satzende mit „-1“ markiert. Des Weiteren wurden die Umlaute ausgeschrieben und „ß“ durch „ss“ ersetzt: Da Catpac nur die englischen Buchstaben kennt und alle anderen Zeichen durch Leerstellen ersetzt, würden dadurch die Wörter an den entsprechenden Stellen in zwei Teile zerlegt.
2. Danach wurde eine Stoppwörterliste mit typischen Stoppwörtern, wie zum Beispiel Artikel oder Personalpronomen, und inhaltlich offensichtlich bedeutungslosen Wörtern aufgebaut.

Bei einer Stoppwörterliste handelt es sich um eine Liste, die alle Wörter enthält, die von der Analyse ausgeschlossen werden. Beispiele für typische Stoppwörter in den Texten der Frankfurter Allgemeinen Zeitung sind „das“, „denen“, „der“, „die“, „ein“, „eine“, „einem“, „einen“, „hat“, „hätten“, „im“ und „in“. Später wurde die Liste um inhaltlich nicht eindeutig zuordenbare oder aber für die Analyse bedeutungslose Wörter ergänzt. Als Beispiel sind für das Wort „Bevölkerung“ im Folgenden zwei Textstellen aufgeführt:

„Die beduinische *Bevölkerung* des Iraks gehört nach den Erkenntnissen Max von Oppenheims, eines der bedeutendsten Kenner der arabischen Wüstennomaden, zu den großen beduinischen Völkern der Schammar und der Aneza (Anaiza).“

„Nördlich des 36. Breitengrades wird eine Schutzzone für die kurdische *Bevölkerung* eingerichtet.“

Das Wort „Bevölkerung“ wurde in die Stoppliste aufgenommen, da es bezüglich unserer Fragestellung nicht eindeutig ist und ihm erst durch die beschreibenden Adjektive „beduinische“ beziehungsweise „kurdische“ inhaltlich eine Bedeutung zuge-

wiesen wird. Diese Adjektive werden auch ohne das Stoppwort „Bevölkerung“ adäquat in die Klassifikationsverfahren einbezogen.

Weitere Beispiele für solche Stoppwörter aus den deutschen Texten sind „Frau“, „Mehrheit“, „Tag“, „Süden“, „Journalisten“, „Unterstützung“, „Land“, „Rolle“, „April“, „Außenminister“ und „Jahr“.

3. Die letzten drei Schritte zur Textvorbereitung waren Arbeiten an den Basistexten selbst, die für die Co-Occurrence-Analyse spezifisch sind. Um verfälschte und inhaltlich unsinnige Ergebnisse zu vermeiden, sahen wir uns gezwungen, Lemmatisierung im Text per Hand vorzunehmen, inhaltlich fest zusammengehörende Wörter (Wortpaare und Phrasen) zusammenzufassen und Synonyme anzugleichen.

Lemmatisierung bezeichnet das Zurückführen der Wörter auf ihre Wortgrundform (das Lemma). Ein Beispiel für Lemmatisierung in unseren deutschen Texten ist die Ersetzung der Wörter „Amerikas“, „Amerikaner“, „Amerikanern“, „amerikanisch“, „amerikanische“, „amerikanischen“ und „amerikanischer“ durch „Amerika“.

4. Nach der sehr umfangreichen und aufwendigen Lemmatisierungsprozedur suchten wir nach inhaltlich fest zusammengehörenden Wortpaaren und Phrasen. Um Wortpaare zu identifizieren, wurden neben der Festlegung von bekannten zusammengehörenden Wortpaaren (wie zum Beispiel „Vereinte Nationen“) erste Clusteranalysen mit Catpac gerechnet, deren Ergebnisse zusätzlich halfen, Wortpaare und Phrasen zu finden. Diese Wortpaare beziehungsweise Phrasen wurden dann im Text zusammengefasst, weil sonst die Gefahr besteht, dass diese Wortkombinationen jeweils eigene Cluster bilden, die für die Inhaltsanalyse an sich wertlos sind. Beispiele für solche Wortkombinationen sind: „United States“, „White House“, „Security Council“, „Al Qaida“, „Al Jazeera“, „Vereinigte Staaten“ und „Vereinte Nationen“.

5. Im letzten Schritt zur Textvorbereitung wurden Synonyme aneinander angeglichen. Ein Beispiel für solch eine Angleichung ist in unserem Fall die Überführung des Wortes „USA“ in das Synonym „Amerika“. Diese Editionsarbeiten wurden allerdings ausschließlich für die Synonyme der häufigsten Wörter ausgeführt.

Nach all diesen Vorarbeiten hatten wir einen Text, der mit Catpac analysiert werden kann. In Abbildung 1 ist ein Textausschnitt mit markierten Textänderungen zu Illustrationszwecken dargestellt.

Abbildung 1: Textbeispiel nach den Editierarbeiten

Aller Rhetorik zum Trotz betrachtet diese Regierung die **Vereintennationen** nur als **muehsames**, letztlich unwesentliches Hindernis auf dem Weg zur Durchsetzung der eigenen Interessen.

-1

Die Tatsache, dass es Frankreich wagen kann, sich dem universalen Anspruch der **Amerika** entgegenzustellen, ist **fuer** Bush nicht nur eine unerhoerte Anmassung, sondern auch Grund zur Rache.

-1

Bush vergisst nicht so leicht, wie er gestern deutlich machte.

-1

4.3 Inhaltsanalyse

Bei einer Inhaltsanalyse mit Catpac wird versucht, Assoziationsmuster zwischen den häufigsten (sinntragenden) Wörtern zu identifizieren und diese dann mit Hilfe von Klassifikationsverfahren darzustellen. Als Klassifikationsverfahren werden Clusteranalysen beziehungsweise in der Zusatzsoftware von Catpac „perceptual maps“, ein multidimensionales Skalierungsverfahren, angeboten.

Die Basis der Analyse bildet ein neuronales Netzwerk, in dem die häufigsten Wörter im Text als Knoten dargestellt werden. Diese Knoten sind durch Relationen verbunden, die durch das gemeinsame Auftreten von Wörtern bestimmt werden. Die Relationen werden gestärkt, wenn zwei Wörter gemeinsam vorkommen, und werden geschwächt, wenn diese Wörter nicht gemeinsam auftreten. Die Idee für diese Vorgehensweise ist der Versuch, die Psychologie des Erinnerns und Vergessens in einem menschlichen Gehirn zu simulieren: Was lange nicht gehört/gelesen wird, wird nach und nach wieder vergessen, und was ständig wiederholt wird, bleibt im Gedächtnis.

Der Bereich, in dem das gemeinsame Vorkommen von Wörtern geprüft wird, kann unterschiedlich definiert werden: als Textfenster fester Länge (zum Beispiel fünf Wörter) oder als logische Einheiten (zum Beispiel Zeitungsartikel). In Catpac hat man die Wahl zwischen drei Gleitoptionen der Textfenster. Eine Option bietet an, dass ein immer gleich großes Fenster über den gesamten Text (ausschließlich den Stoppwörtern) geschoben wird und hier die Verbindungen der häufigsten Wörter ermittelt werden. Eine zweite Option lässt das Fenster nur über die häufigsten Wörter gleiten. Die restlichen Wörter finden bei der Analyse keine weitere Beachtung. Bei einer dritten Option kann der Anwender die Texteinheiten festlegen (zum Beispiel Sätze oder Zeitungsartikel). Diese Einheiten wer-

den als (unterschiedlich große) Fenster behandelt. Das variabel große Fenster gleitet in diesem Fall von Texteinheit zu Texteinheit.

Abhängig von der Wahl der Gleitoption wird ein Fenster der Größe n über die Wörter 1 bis n des Textes gelegt⁵. Kommt ein Wort im Fenster vor, wird der Knoten, der durch das Wort definiert ist, aktiviert. Die Knoten⁶, die gleichzeitig in einem Fenster aktiv sind, werden virtuell miteinander verbunden. Dann wird das Fenster um ein Wort nach rechts geschoben, das heißt, es umfasst die Wörter 2 bis $n+1$ ⁷. Nun werden neue Knoten im Netzwerk aktiviert und deren Relationen erneut berechnet. Das Fenster wird so wortweise über den Text gezogen, bis das Textende erreicht ist. Die Verbindung zwischen Wörtern, die wiederholt gemeinsam in einem Fenster vorkommen, wird im Verlauf der Analyse stärker und zwischen wiederholt nicht aktivierten Wörtern entsprechend schwächer. Die so berechneten Ähnlichkeitskoeffizienten bilden die Basis für die daran anknüpfende Clusteranalyse und gegebenenfalls die MDS.

Wir haben uns bei dem folgenden Anwendungsbeispiel aus methodischen Erwägungen für die dritte Gleitoption, bei der der Anwender die Texteinheiten bestimmt und die Fenster entsprechend unterschiedlich groß sind, entschieden: Die Option des gleitenden, immer gleich großen Fensters erschien uns problematisch, da kein plausibler Grund erkennbar war, warum ein Fenster zum Beispiel fünf oder sieben Wörter lang sein soll. Die zweite Option erschien uns ebenfalls wenig sinnvoll, da das Gleiten eines gleich großen Fensters über nur die häufigsten Wörter, das heißt, alle Relationen zu den weniger häufigen Wörtern werden außer Acht gelassen, zusätzlich einen vermeidbaren Informationsverlust erwarten ließ.

Als Texteinheiten haben wir Sätze gewählt, da dies semantisch logische Einheiten in Zeitungsartikeln sind. In unserem ersten Versuch hatten wir uns für Artikel als Texteinheiten entschieden, es hat sich jedoch schnell gezeigt, dass diese Einheiten für die Analyse zu groß gewählt waren: Ein Großteil der für unsere Analyse wichtigen Wörter wurde gemeinsam in einem Cluster zusammengeführt und der Cluster hatte damit keinerlei Aussagekraft bezüglich unserer Fragestellung. Der Grund hierfür war, dass nahezu alle inhaltlich interessanten Wörter gemeinsam in den meisten Artikeln auftraten.

Zusätzlich wurde festgelegt, dass die 25 häufigsten Wörter in die Analyse eingehen, weil sich diese Wörter als die inhaltlich aussagekräftigsten herauskristallisiert haben. Alle

5 Die Zahl der Wörter (n), die ein Fenster umfassen soll, wird vom Anwender bestimmt.

6 Die Zahl der in die Analyse eingehenden Wörter wird vom Anwender festgelegt.

7 Ob das Fenster ein oder mehrere Wörter nach rechts verschoben wird, wird vom Anwender definiert.

weiteren Wörter waren entweder als Stoppwörter zu behandeln oder hatten nur sehr geringe Häufigkeiten.

Auf die Bestimmung der Assoziationsmaße kann durch Einstellungen in den sogenannten „Learning Parameters“ Einfluss genommen werden. Hier wird festgelegt, wie schnell innerhalb eines Netzwerks die Knoten (beziehungsweise deren Verbindungen) gestärkt beziehungsweise wieder geschwächt werden. Je kleiner der Wert für die „Learning rate“ ist, desto langsamer lernt das Programm (siehe auch Doerfel 1994:11). Da uns keine Argumente für die Regulierung der „Learning rate“ nach oben oder nach unten vorlagen, entschieden wir uns für die vom Programm voreingestellte „Learning rate“ von 0.01.

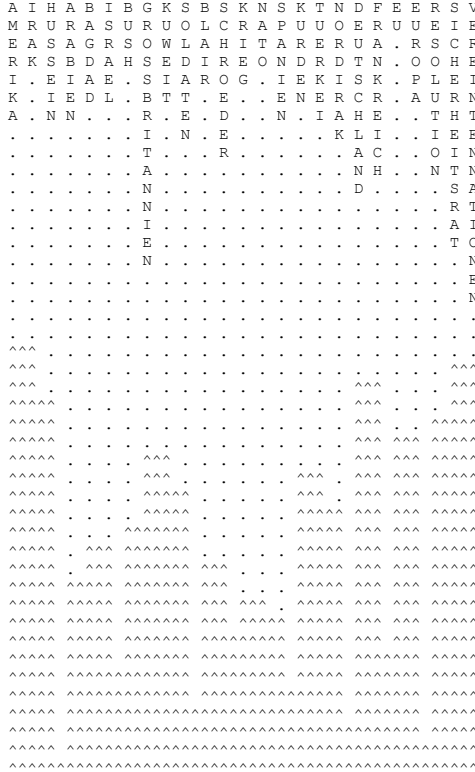
Die Basis für die hierarchische Clusteranalyse bilden die im Netzwerk berechneten Ähnlichkeitskoeffizienten. Bei den in diesem Artikel beschriebenen Analysen kam das Ward'sche Verfahren zum Einsatz, das die Varianz innerhalb einer Gruppe möglichst gering hält und möglichst gleichmäßig besetzte Gruppen bildet.

4.4 Einige Ergebnisse

Wie oben beschrieben, gleitet ein virtuelles Fenster über die vier vorbereiteten Textbasen und Catpac stellt als Ergebnis einer hierarchischen Clusteranalyse das berechnete Assoziationsmuster der jeweiligen Texte dar. Bei den von Catpac angebotenen Dendrogrammen, mit deren Hilfe das Ergebnis der Analyse dargestellt wird, handelt es sich nicht um die üblicherweise erzeugten Dendrogramme, sondern um Diagramme, die in anderen Programmen, zum Beispiel in SPSS, als Icicle-Plots bezeichnet werden. Diesen Diagrammen kann entnommen werden, in welcher Reihenfolge die Wörter zusammengefasst wurden und wieviele Cluster es gibt, allerdings nicht, wie in regulären Dendrogrammen, welche Distanzen zwischen den Clustern bestehen.

Da die Diagramme schlecht lesbar sind (siehe Abbildung 2), haben wir die Ergebnisse grafisch nachbereitet. Abbildung 3 zeigt solch ein neu erstelltes Diagramm. Auf der linken Seite der Grafik sind die Wörter angegeben. Das Zusammenfassen zweier Wörter zu einem Aggregat ist dadurch gekennzeichnet, dass die Linie, die von den beiden Wörtern ausgeht, am Ende zu einem Balken geschlossen wird. Jeder Balken steht für eine Aggregationsstufe, das heißt, die Grafik enthält genau einen Balken weniger als Wörter. Die Grafik ist von rechts nach links zu lesen. Je länger der Balken ist, desto früher wurden die Wörter zusammengeführt. Der längste Balken zeigt, welche Wörter auf der ersten Aggregationsstufe zusammengefasst wurden (in Abbildung 3 Amerika und Irak), das heißt, es handelt sich hier um die am engsten miteinander verbundenen Wörter. Auf der nächsten Aggregationsstufe werden Sicherheitsrat und Vereinte Nationen zusammengefasst, gekennzeichnet durch den zweitlängsten Balken.

Abbildung 2: Iccle-Plot für die Texte der Frankfurter Allgemeinen Zeitung zu Kriegsbeginn im Originalformat

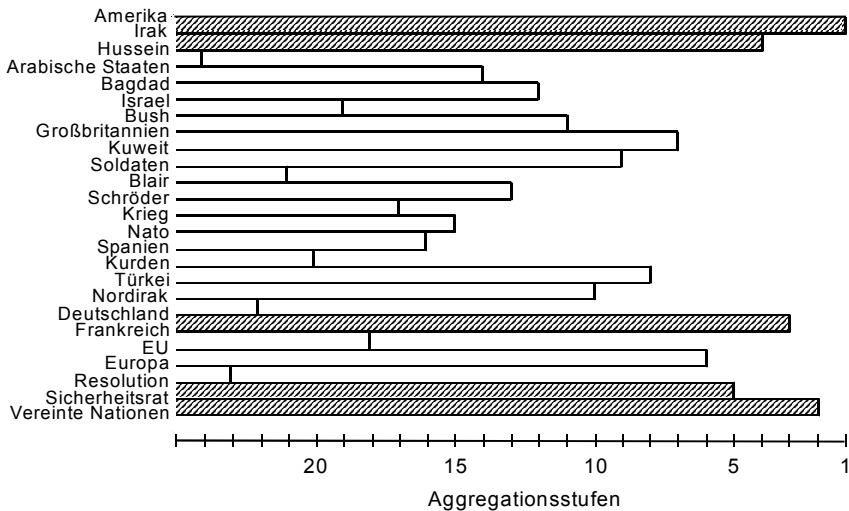


Auf der dritten Aggregationsstufe werden Deutschland und Frankreich zusammengeführt und auf der vierten Stufe wird das Wort „Hussein“ den ersten beiden Wörtern zugefügt. Diese Aggregationen werden so lange durchgeführt, bis alle Wörter miteinander verbunden sind. Direkt nebeneinanderliegende Balken, die relativ kurz nacheinander zusammengefügt werden, werden später als Cluster interpretiert. Im Gegensatz zu Dendrogrammen sind aus diesen Grafiken allerdings keine Abstände zwischen Wörtern/Clustern abzulesen. Die in unserer Analyse interpretierten Cluster sind zur besseren Lesbarkeit schraffiert.

Die für die Texte der New York Times und die Texte der Frankfurter Allgemeinen Zeitung jeweils für Kriegsbeginn und -ende in Catpac errechneten Clusterlösungen (siehe Abbil-

dung 3 bis 6) werden in diesem Abschnitt anhand der drei zuerst zusammengeführten Cluster deskriptiv interpretiert. Diese Cluster weisen die stärkste Verbindung auf und sind damit inhaltlich die bedeutsamsten. Aus Gründen der besseren Handhabung der Vergleiche zwischen den verschiedenen Zeitpunkten der Analyse und den zwei Zeitungen werden den Clustern bei Bedarf Labels zugeordnet. Zur Vermeidung willkürlicher Interpretationen werden Textstellen, in denen die betreffenden Wörter sehr nahe beieinander auftreten, gesucht und quasi als Interpretationshilfe beziehungsweise -legitimation herangezogen.

Abbildung 3: Clusteranalyse der Texte der Frankfurter Allgemeinen Zeitung zu Kriegsbeginn (Texteinheit = 1 Satz)



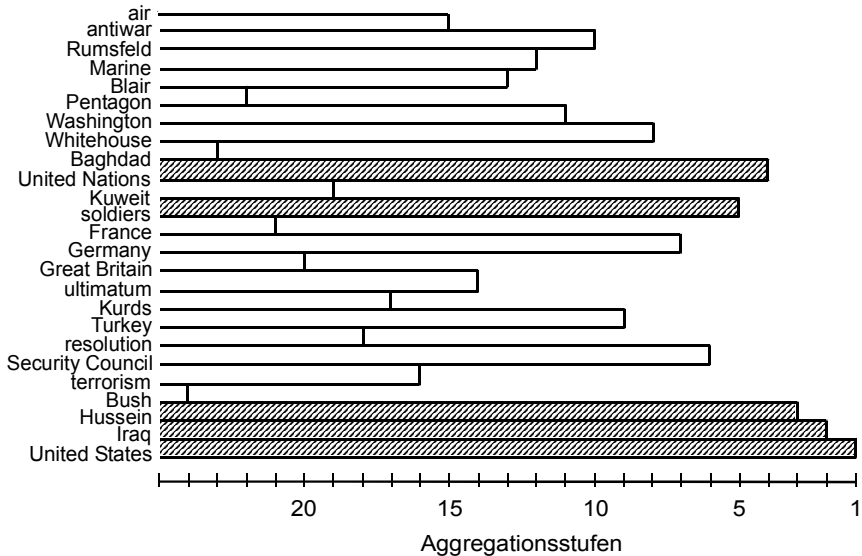
Im Diagramm zur Berichterstattung in der Frankfurter Allgemeinen Zeitung am Anfang des Krieges sind insgesamt neun Cluster zu erkennen. Die drei Cluster, die auf den ersten Aggregationsstufen zusammengeführt werden, sind »Amerika, Irak, Hussein«, »Sicherheitsrat, Vereinte Nationen, Resolution« und »Frankreich, Deutschland«. Eine Original-Textstelle für jeden der drei Cluster ist als Beispiel für die Berichterstattung der Frankfurter Allgemeinen Zeitung zu Kriegsbeginn im Folgenden aufgeführt. Die die Cluster definierenden Wörter sind in den Texten kursiv dargestellt.

„Der *Irak* weist Bushs Ultimatum zurück. Der Revolutionäre Kommandant, das höchste, von *Hussein* angeführte Entscheidungsgremium des *Irak*, hat am Dienstag das *amerikanische* Ultimatum an *Hussein* und seine beiden Söhne abgelehnt.“

„Der *UN-Sicherheitsrat* habe mit seiner Unfähigkeit, eine gemeinsame Haltung zu der *Resolution* zu finden, abermals gezeigt, dass er seiner Verantwortung nicht gerecht werde. Rußland sieht in den bisherigen *Resolutionen* der *Vereinten Nationen* "keine rechtliche Grundlage" für einen Militärschlag gegen den *Irak*, sagte Außenminister Iwanow am Montag abend in Reaktion auf die Entscheidung Großbritanniens und Amerikas.“

„Unter anderen werden die Außenminister *Frankreichs* und *Deutschlands*, de Villepin und Fischer, erwartet. Was der Rat beraten soll, war in New York unklar. *Deutschland* und *Frankreich* kritisierten Bushs Entscheidung.“

Abbildung 4: Clusteranalyse der Texte der New York Times zu Kriegsbeginn (Texteinheit = 1 Satz)



Entsprechend dem Inhalt der Textstellen erhielt der Cluster »Amerika, Irak, Hussein« das Label ‚Hauptkonfliktparteien‘, der Cluster »Resolution, Sicherheitsrat, Vereinte Nationen« steht für ‚Kriegsvermeidung‘ und der dritte Cluster bekam das Label ‚Kriegsgegner‘.

Bei der Interpretation der Clusterlösung zur anfänglichen Kriegsberichterstattung in den Vereinigten Staaten wurde gleichermaßen vorgegangen. Bei den drei Clustern, die hier auf den obersten Aggregationsstufen zusammengeführt werden, handelt es sich um »United States, Iraq, Hussein, Bush«, »Baghdad, United Nations« und »Kuwait, Soldiers«. Nachstehend ist wiederum je ein Textbeispiel aus der New York Times für jeden der drei Cluster aufgeführt:

„Even if Saddam *Hussein* leaves Iraq within 48 hours, as President *Bush* demanded, allied forces plan to move north into *Iraqi* territory, *American* officials said.”

„About 60 *United Nations* weapons inspectors, along with about 100 staff members, were expected to leave *Baghdad* on Tuesday morning.”

„This afternoon, *soldiers* of the Third Infantry Division's First Brigade Combat Team began packing up and dismantling parts of a mobile command center in the *Kuwaiti* desert.”

Vor allem sind die zwei zuletzt genannten Cluster interessant. Im Gegensatz zur Frankfurter Allgemeinen Zeitung spielt bei der New York Times die konkrete operative Ebene, das heißt die konkrete Umsetzung der Intervention, eine größere Rolle als die abstrakt globalpolitisch geführte Diskussion um deren Legitimation. Entsprechend stehen sich die an zweiter Stelle zusammengeführten Cluster »Sicherheitsrat, Vereinte Nationen, Resolution« und »Baghdad, United Nations« der beiden Zeitungen gegenüber. Während sich die Frankfurter Allgemeine Zeitung noch mit der vermeintlichen Kriegsvermeidung durch Hinweise auf die fehlende politische völkerrechtliche Legitimation für einen Militärschlag durch die UN beschäftigt, beschäftigt sich die New York Times bereits mit dem Abzug der UN-Inspektoren und damit dem ‚Spielerwechsel‘ auf operativer Ebene. Eine ähnliche Differenz ist auch beim Vergleich der an dritter Stelle zusammengeführten Cluster festzustellen. Hier steht der Cluster »Kuweit, Soldiers« im Prinzip bereits für den langsam beginnenden Krieg. Inhaltlich steht der Cluster dafür, dass sich die in Kuwait stationierten Soldaten für den Abmarsch, das heißt den Bodenkrieg im Irak, bereit machen. Ein inhaltlich ähnlicher Cluster wird im Fall der Frankfurter Allgemeinen Zeitung erst als fünfter Cluster zusammengeführt, während der „dritte Platz“ der Kriegsgegnerschaft von Frankreich und Deutschland vorbehalten ist, die in der amerikanischen Zeitung wiederum nur einen der „hinteren Plätze“ belegt.

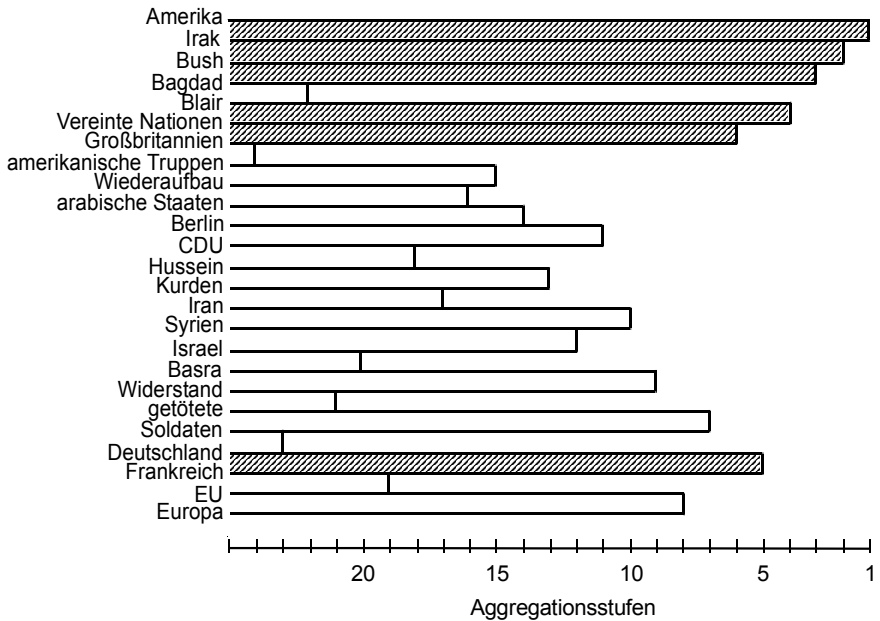
Der in der New York Times als erstes zusammengeführte Cluster »United States, Iraq, Hussein, Bush« ist dem an erster Stelle befindlichen Cluster der Frankfurter Allgemeinen Zeitung »Amerika, Irak, Hussein« inhaltlich sehr ähnlich und erhält daher ebenfalls das Label ‚Hauptkonfliktparteien‘. Dass „Bush“ im Fall der Frankfurter Allgemeinen Zeitung in diesem Cluster keinen Platz gefunden hat, ist mit Blick auf die Gleichsetzung des

Hauptrepräsentanten eines Landes mit seinem Land durch die ausländische Presse plausibel und soll an dieser Stelle als mögliche Begründung genügen.

Fazit der Betrachtung der Texte zu Kriegsbeginn ist, dass das tatsächliche Ausmaß der Kriegsgefahr in der New York Times sehr viel deutlicher zum Ausdruck kam.

Im Gegensatz zu den für Beginn der Kriegsberichterstattung gebildeten Cluster weisen die drei für die Berichterstattung am Ende des Krieges gebildeten Cluster inhaltlich erwartungsgemäß andere Ausprägungen auf (siehe Abbildung 5 und 6).

Abbildung 5: Clusteranalyse der Texte der Frankfurter Allgemeinen Zeitung zu Kriegsende (Texteinheit = 1 Satz)



Bei den aus dem Text der Frankfurter Allgemeinen Zeitung gebildeten Clustern zum Kriegsende handelt es sich um »Amerika, Irak, Bush, Bagdad«, »Blair, Vereinte Nationen, Großbritannien« und »Deutschland, Frankreich«. Es folgt zunächst wieder zu Interpretationszwecken ein konkreter Textauszug je gebildeten Cluster:

„Das *irakische* Volk ist vollkommen dazu in der Lage, den *Irak* selbst zu regieren“, sagt Präsident *Bush*. Und: „Die *Amerikaner* werden beim Übergang helfen.“⁶⁴

aber auch

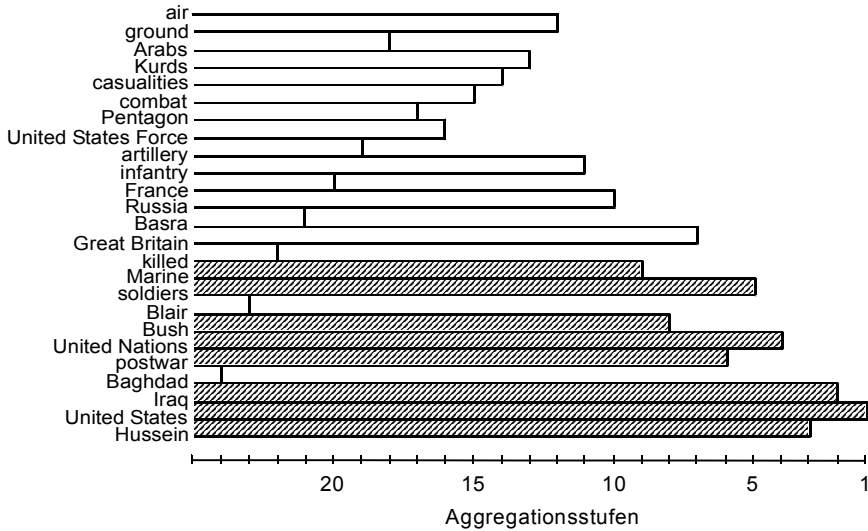
„Der *irakische* Informationsminister bestreitet den Vorstoß der *Amerikaner* nach *Bagdad*.“

„Auf Fragen des Oppositionsführers Duncan Smith und des liberaldemokratischen Parteivorsitzenden Kennedy im Parlament hatte *Blair* vorige Woche ausweichend geantwortet, der Irak solle "weder von der Koalition noch von den *Vereinten Nationen* regiert werden, sondern vom irakischen Volk". In *Großbritannien* hieß es, *Blair* hoffe, er könne *Bush* zu einer Änderung des Plans überreden, in der irakischen Übergangsregierung zwanzig amerikanische "Minister" einzusetzen.“

„War es im Falle Irak tatsächlich die einzige Lösung? Wir hätten es verhindern können, wenn *Deutsche* und *Franzosen* bei der militärischen Drohung gegen Saddam Hussein mitgemacht hätten. Wenn wir alle Schulter an Schulter gestanden hätten, wäre der Druck auf Bagdad größer gewesen.“

Der erste Cluster wurde mit dem ‚Nachkriegsphase_USA‘-Label versehen. Inhaltlich sind hier jedoch zwei Schwerpunkte zu unterscheiden: zum einen die Planung, speziell der USA, für den Nachkriegs-Irak und zum anderen letzte Meldungen über Kriegshandlungen in Bagdad. Im zweiten Cluster kommt die Diskussion um den Nachkriegs-Irak in Großbritannien zum Ausdruck und er erhält das Label ‚Nachkriegsphase_UK‘. Die inhaltliche Auseinandersetzung um die ‚Kriegsgegnerschaft‘ wird durch den Cluster »Frankreich, Deutschland« repräsentiert. In den Texten wird nun die Kriegsgegnerschaft der beiden Länder relativiert.

Abbildung 6: Clusteranalyse der Texte der New York Times zu Kriegsende (Texteinheit = 1 Satz)



Im Gegensatz zur Frankfurter Allgemeinen Zeitung fokussiert der erste Cluster »United States, Iraq, Baghdad, Hussein« bei der New York Times einen einzigen Schwerpunkt und zwar die letzten Meldungen über Kriegshandlungen in Bagdad. Die zusammengeführten Wörter in diesem Cluster und in dem ‚Nachkriegsphase‘-Cluster der Frankfurter Allgemeinen Zeitung sind fast identisch, nur „Bush“ und „Hussein“ sind vertauscht. Trotzdem handelt es sich bei dem zuerst zusammengeführten Cluster der New York Times nicht um einen ‚Nachkriegsphase‘-Cluster, was die Notwendigkeit, die Wörter auch in ihrem vollständigen Textumfeld zu betrachten, noch einmal verdeutlicht. An zweiter Stelle ist bei der New York Times ein eindeutiger ‚Nachkriegsphase‘-Cluster auszumachen. Dieser Cluster »Bush, United Nations, Blair, Postwar« veranschaulicht die zum Teil personalisierten potentiellen Entscheidungsträger für die Nachkriegszeit und stellt Großbritannien und die USA im Gegensatz zur Frankfurter Allgemeinen Zeitung als Einheit dar. Der dritte zusammengeführte Cluster »Soldiers, Marine, Killed« steht abermals für eine reine Kriegsberichterstattung, die vermutlich für die Presseberichterstattung der kriegsführenden Nation typisch ist. Beispiele für Textstellen sind folgende:

„As *American* infantry troops encircle *Baghdad* and make thrusts into the city itself, top *Iraqi* military commanders are apparently still conveying positive messages to the younger Mr. *Hussein*, who was appointed leader of the security forces by his father before the war began and who is reputed to be a cunning and brutal officer.”

„He envisioned no central *postwar* role for the *United Nations*, a subject that will be on the agenda when President *Bush* meets Prime Minister Tony *Blair* of Britain in Northern Ireland on Monday.”

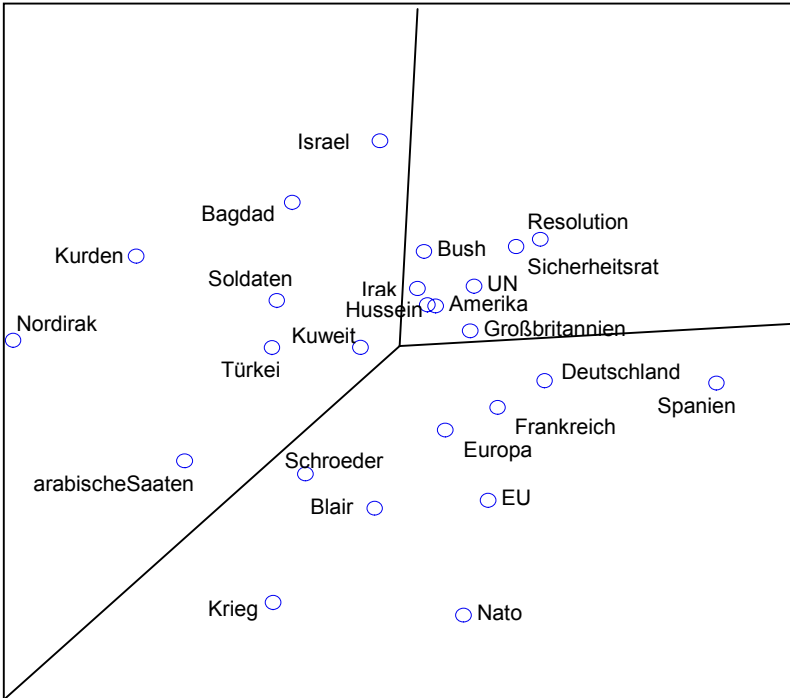
„By contrast, he said, the *marines* and *soldiers* now in combat in Iraq "have a pretty good feel for where the enemy is and they have very, very precise weapons to attack him, and so that makes a huge difference. ""” und “Three British *soldiers* were *killed*.”

Es lässt sich konstatieren, dass bei der deutschen Berichterstattung der Frankfurter Allgemeinen Zeitung am Anfang sowie auch am Ende im Gegensatz zur amerikanischen Berichterstattung der New York Times mehr Betonung auf den politischen Charakter des Krieges gelegt wurde. Es fällt auf, dass die Vereinten Nationen bei der New York Times im Gegensatz zur Frankfurter Allgemeinen Zeitung niemals im ersten Cluster auftraten und die Kriegsgegnerallianz Frankreich/Deutschland in der Berichterstattung der New York Times zwar zur Kenntnis genommen wurde, aber nicht wirklich zentrales Thema war. Außerdem wurde bei der Berichterstattung der New York Times im Gegensatz zur Frankfurter Allgemeinen Zeitung bei der Diskussion um die Ausgestaltung des Nachkriegs-Iraks nicht zwischen Großbritanniens Meinung und der der Vereinten Staaten unterschieden. Die Gründe für die hier nicht erschöpfend diskutierten Unterschiede der Profile könnten in national divergierenden Interessen und Betroffenheit zwischen beiden Nationalstaaten zu finden sein. Eine fundierte Einschätzung der unterschiedlichen Profile der beiden Zeitungen kann eine Inhaltsanalyse ohne theoretische Vorüberlegungen jedoch nicht bieten.

Im Anschluss an die Clusteranalysen wurde mit dem Zusatzmodul von Catpac, Thoughtview, eine multidimensionale Skalierung (MDS) gerechnet. Es zeigte sich dabei, dass das Programm zwar eine grafische Lösung darstellt, aber keinerlei statistische Informationen zur Prüfung der Güte der Lösung, wie zum Beispiel den Stress-Wert oder ein Shepard-Diagramm, anbietet. Wir haben uns deshalb entschieden, die multidimensionale Skalierung mit dem Statistikprogramm Systat zu rechnen. Da sich bei den ersten Analysen zeigte, dass einzelne Werte der Ähnlichkeitsmatrix extreme Werte annahmen, wurde die Matrix standardisiert. Dazu wurde die Kovarianzmatrix aus Catpac in eine Korrelationsmatrix umgewandelt, die als Basis für die lineare MDS diente. An dieser Stelle soll exemplarisch nur auf ein Ergebnis eingegangen werden, das für den Text der Frankfurter Allgemeinen Zeitung zu Kriegsbeginn gerechnet wurde. Der Stress-Wert betrug bei dieser Analyse einen akzeptablen Wert von 0.24 (siehe Bewertungskriterien in Borg/Groenen 1997). In Abbildung 7 zeigen sich drei Regionen. Die Region auf der linken Seite (,arabi-

sche Staaten') umfasst die arabischen Staaten, die Türkei und Israel. Rechts unten dagegen befinden sich die europäischen Staaten. Rechts oben und sehr eng miteinander verbunden ist eine Region der kriegsführenden Parteien (Amerika, Großbritannien, Hussein, Bush), die aber auch eng mit der UN und der Friedensresolution assoziiert ist, zu erkennen. Hier spiegelt sich die Diskussion in der Frankfurter Allgemeinen Zeitung um die Kriegsvermeidung wieder. Das Wort „Krieg“ liegt ebenso am Rande des Diagramms wie „Nato“. Dies könnte darauf hindeuten, dass die beiden Wörter im zweidimensionalen Raum keine systematische Beziehung zu den anderen Themen im Text haben. So wird das Wort „Nato“ sehr unsystematisch verwendet, zum Beispiel geht es um Nato-Partner, Nato-Beitritt der Osteuropäer, Rolle der Nato zur Kriegsvermeidung und um Aufklärungsflugzeuge der Nato in der Türkei.

Abbildung 7: Multidimensionale Skalierung der Texte der Frankfurter Allgemeinen Zeitung zu Kriegsbeginn



5. Diskussion

5.1 Probleme bei der Verwendung von Catpac

Obwohl Catpac bereits in verschiedenen Projekten eingesetzt wurde, hat sich nach, beziehungsweise während des Rechnens mit unserem Anwendungsbeispiel gezeigt, dass der dort praktisch umgesetzte Ansatz noch einer Weiterentwicklung bedarf. Zwar haben wir mit dem Einsatz von neuronalen Netzen, der Clusteranalyse und der anschließenden MDS Themen beziehungsweise Wortcluster identifiziert und interpretiert, jedoch bieten Catpac und Thoughtview keinerlei Hilfsmittel, die dem Anwender eine gesicherte Einordnung und Bewertung der Ergebnisse erlauben. Die Programme bieten keine statistischen Informationen, wie zum Beispiel Stress-Maße oder Shepard-Diagramme, und anstelle eines vollwertigen Dendrogramms, aus dem Abstände zwischen den Clustern zu sehen wären, wird lediglich ein Icicle-Plot angeboten. Hier ist zwar die Rangfolge der Fusionen ersichtlich (was zuerst und was zuletzt zusammengefügt wird), aber es ist nicht ersichtlich, wie groß der Abstand zwischen den Aggregaten ist. Auch gibt es keine Fusionswertetabelle, aus der diese Abstände abzulesen wären. Die Bestimmung der Zahl der Cluster und damit auch der Inhalt der Cluster unterliegt damit einer gewissen Willkür, was sich auf eine Interpretation der Ergebnisse nachhaltig auswirkt.

Weitere Probleme zeigten sich bei der Auswahl der verschiedenen Programmoptionen. Es gibt bisher keine Erfahrungen, wie die Optionen, die das Lernen und Vergessen im Netzwerk steuern, gesetzt werden sollten. Beispielsweise simuliert ein hoher Wert bei der „Learning rate“ schnelleres und ein niedriger Wert langsames Lernen, aber welcher Wert nun tatsächlich sinnvoll ist, dazu gibt es keine Erfahrungen. Problematisch ist dieser Umstand bei der „Learning rate“ insofern, da in Abhängigkeit von dieser Einstellung unterschiedliche Ähnlichkeiten zwischen zwei Wörtern und damit unterschiedliche Cluster berechnet werden. Dem Entwickler ist diese Problematik nicht neu und so schreibt er selbst: „No one knows the optimum rate, or even if there is an optimum rate, so feel free to experiment“ (Woelfel 1998:25).

5.2 Probleme bei der Anwendung der Co-Occurrence-Analyse

Hat man sich für die Anwendung einer Co-Occurrence-Analyse entschieden, ist bereits die Wahl der geeigneten Software entscheidend. Die Softwareprogramme zur Co-Occurrence-Analyse unterscheiden sich stark in den verwendeten Klassifikationsverfahren (Ähnlichkeitsmaße, Clusterverfahren, multidimensionale Skalierung). Dies führt zu grundsätzlich unterschiedlichen Ergebnissen und schließt einen Vergleich der Ergebnisse im Vorhinein nahezu aus. Problematisch ist dieser Umstand insofern, dass die Wahl des „geeigneten“ Verfahrens vor den ersten Analysen meist relativ schwierig ist und jedes

Programm in der Regel eine andere Art von Textvorbereitung erfordert, das heißt, ein schneller Wechsel zwischen den Programmen ist kaum möglich. Man hat jedoch immer die Option, nach Erstellen der Ähnlichkeitsmatrix in einem Textanalyseprogramm, alle weiteren statistischen Analysen in entsprechenden Statistikprogrammen zu rechnen, wie wir es am Beispiel der MDS gezeigt haben.

Unabhängig von der verwendeten Software sind die Arbeiten zur Vorbereitung der Texte sehr aufwendig: Zwar wird die Liste der häufigsten Wörter in der Regel automatisch generiert und man kann problemlos eine Stoppwörterliste erstellen, aber es fehlt an technischer Unterstützung bei der Lemmatisierung. Derzeit üblich ist das manuelle Vorgehen per „Suchen und Ersetzen“.

Des Weiteren war nach Betrachtung der Texte eine aufwendige Bearbeitung der Synonyme nötig, da es in der Regel keinen Sinn macht, gleichbedeutende Wörter unterschiedlich zu behandeln (zum Beispiel „Vereinte Nationen“ und „UN“). Wie weit man bei der Definition von Synonymen geht, ist abhängig von der Fragestellung. Im einfachsten Fall werden nur wirklich gleiche Begriffe zusammengefasst (zum Beispiel „Vereinte Nationen“ und „UN“). Man kann sich aber auch dafür entscheiden, die Definition von Synonymen zu verfeinern und Wörter mit ähnlicher Bedeutung zusammenfassen, zum Beispiel könnte das Synonym „Naher Osten“ aus den Wörtern „Israel“, „Libanon“ und „Syrien“ gebildet werden. Im letztgenannten Fall würde man allerdings einer Diktionärskonstruktion schon sehr nahe kommen.

Ferner fehlen unseres Erachtens konzeptionelle Überlegungen zur Festlegung der Länge einer Texteinheit, da die Größe einer Texteinheit die Ergebnisse der Klassifikation stark beeinflusst. Es ist plausibel, dass es zu verschiedenen Ergebnissen führen muss, wenn man das gemeinsame Auftreten von Schlüsselwörtern innerhalb von fünf Wörtern oder innerhalb eines Satzes oder innerhalb eines Artikels untersucht (siehe Danowski 1993:209 und Galliker/Hermann 2003:99). Von den zwei üblicherweise praktizierten Vorgehensweisen, Texteinheiten zu definieren – entweder orientiert sich die Wahl der Texteinheit an der syntaktischen Struktur des Textes (zum Beispiel Sätze oder Artikel) oder man arbeitet mit Texteinheiten, die alle die gleiche Länge haben (zum Beispiel fünf Wörter) – erschien uns jedoch die erstgenannte einleuchtender. Eine Begründung für die Verwendung von gleich langen Texteinheiten ist schwierig. In der Literatur finden sich hierzu Hinweise wie zum Beispiel „A window of five words was selected because past research had shown that it would be sufficiently wide to accommodate the subject-verb-object syntax in English and not so wide as to allow words that are not semantically unrelated to appear to be related...“ (Salisbury 2001:71). Dieses Argument erschien uns zunächst plausibel, allerdings konnten wir diesen Sachverhalt bei unseren englischsprachigen Texten aus der New York

Times nicht bestätigen. Aber auch die Orientierung an der syntaktischen Struktur war problematisch. Bezogen auf unsere Fragestellung und die Textart erschien uns der Zeitungsartikel auf den ersten Blick eine sinnvolle Einheit zu sein. Bei Durchführung der Analyse stellten wir jedoch fest, dass die Definition von so großen Einheiten dazu führt, dass häufig alle zentralen Begriffe gemeinsam auftreten und keine klare Clusterstruktur gefunden werden kann.

Offen bleibt auch der Einfluss der Textlänge, der Häufigkeit der Wörter und der Häufigkeit des gemeinsamen Auftretens von Wörtern auf die Ergebnisse, dem wir nicht nachgegangen sind. Es ist aber zu erwarten, dass bei größeren Texten, in denen die ausgewählten Stichwörter häufig gemeinsam vorkommen, die Ergebnisse der neuronalen Netzwerkanalyse beziehungsweise einer einfachen Co-Occurrence-Analyse klarer werden.

6. Fazit

Abschließend lässt sich konstatieren, dass es zum einen inhaltliche Probleme bezüglich der Methodik des Co-Occurrence-Ansatzes gibt und zum anderen der am häufigsten genannte Vorteil, der des geringeren Aufwandes, da kein a priori Diktionär erstellt werden muss, mit Blick auf die unerlässlichen Textvorbereitungsmaßnahmen nicht zutrifft. Unabhängig von der Frage, ob Theorien als Entscheidungshilfen zur Interpretation von Vorgängen in der Gesellschaft überhaupt ersetzt werden können oder ob man durch offene Herangehensweisen an jegliches Datenmaterial neue Theorien überhaupt erst entwickelt, bietet der Co-Occurrence-Ansatz, den wir hier stellvertretend für die explorative Herangehensweise einem Praxistest unterzogen haben, zum jetzigen Zeitpunkt zu wenige Entscheidungshilfen zur klaren Strukturierung beziehungsweise zur Interpretation der Ergebnisse.

Unserer Einschätzung nach ist die Anwendung des Co-Occurrence-Ansatzes eher als ergänzendes Verfahren zu diktionsbasierten Ansätzen zu empfehlen. Mit Hilfe der gefundenen Assoziationsmuster kann die Generierung von Hypothesen und das darauf aufbauende Kategoriensystem bei diktionsbasierten Verfahren vervollständigt werden. Durch die Hinzunahme der mit Co-Occurrence-Analysen erzielten Ergebnisse finden auch die Textinhalte Berücksichtigung, die von den aus der Theorie abgeleiteten Kategorien abweichen. Damit kann zwar keine Zeitersparnis erzielt werden, jedoch möglicherweise eine vollständigere Analyse von Textinhalten.

Um den Co-Occurrence-Ansatz als exploratives Analysewerkzeug zu verbessern, sehen wir Entwicklungsbedarf in mindestens drei Bereichen. Automatische Lemmatisierung, wie sie in der Computerlinguistik bereits eingesetzt wird, würde den Arbeitsaufwand bei

der Analyse deutlich reduzieren. Um Anhaltspunkte für die Wahl der Länge von Texteinheiten zu erhalten, erscheint die Auseinandersetzung mit der inhaltlichen und statistischen Bedeutung der Länge von Texteinheiten notwendig. Anzugehen ist auch die Frage, ob neuronale Netze zur Berechnung der Ähnlichkeiten im Vergleich zu einfachen Co-Occurrence-Maßen, wie zum Beispiel dem Jaccard-Ähnlichkeitsmaß, die Ergebnisse der Co-Occurrence-Analyse wirklich verbessern. Hier fehlen derzeit noch Vergleichsdaten.

Literatur

- Alexa, M., 1997: Computer-assisted Text Analysis Methodology in the Social Sciences. Mannheim: ZUMA.
http://www.gesis.org/Publikationen/Berichte/ZUMA_Arbeitsberichte/97/97_07abs.htm.
- Borg, I./Groenen P., 1997: Modern Multidimensional Scaling. Theory and Applications. New York: Springer.
- Brier, A., 2003: Analysis of Joint Frequencies of Words in a Text. User notes for Hamlet for Windows. <http://www.soton.ac.uk/~apb/hamlet95.pdf>.
- Danowski, J.A., 1993: Network Analysis of Message Content. S.197-222 in: G. A. Barnett/W. D. Richards: Progress in communication sciences. Norwood, NJ: Ablex.
- Doerfel, M. L., 1994: The 1992 Presidential Debates: A new Approach to Content Analysis. Paper prepared for presentation at the annual meeting of the Speech Communication Association, New Orleans, LA.
- Freeman, C.A./Barnett, G.A., 1994: An Alternative Approach to Using Interpretative Theory to Examine Corporate Messages and Organizational Culture. S.60-73 in: L. Thayer: Organization-Communication: Emerging perspectives IV. Norwood, NJ: Ablex.
- Galliker, M./Hermann, J., 2003: Medienpsychologische Methoden: Inhaltsanalyse elektronisch gespeicherter Massendaten der internationalen Presse. Zeitschrift für Medienpsychologie 3: 98-105.
- Gerbner, G./Holsti, O. R./Krippendorf, K./Paisley, W. J./Stone, P. J., 1969: The Analysis of Communication Content: Developments in Scientific Theory and Computer Technology. New York: John Wiley and Sons.
- Hogenraad, R./McKenzie, D.P./Péladeau, N., 2003: Force and Influence in Content Analysis: The Production of New Social Knowledge. Quality & Quantity 221-238.

Iker, H.P./Harway, N.I., 1969: A Computer Systems Approach Toward the Recognition and Analysis of Content. S.381-405 in: G. Gerbner: The Analysis of Communication Content. New York: Wiley & Sons.

Mohler, P.Ph./Züll, C., 2001: TEXTPACK User's Guide. Mannheim: ZUMA.

Rosen, D./Woelfel J.K./Krikorian D./Barnett G.A., 2003: Procedures for Analyses of Online Communities. <http://www.ascusc.org/jcmc/vol8/issue4/rosen.html>.

Salisbury, J.G.T., 2001: Using Neural Networks to Assess Corporate Image. S.65-86 in: M. D. West: Applications of Computer Content Analysis. Norwood, NJ: Ablex.

Schrodt, P.A./Davis, S.G./Weddle, J.L., 1994: Political Science: KEDS - A Program for the Machine Coding of Event Data. Social Science Computing Review 4: 561-587.

Sherblom, J.C./Reinsch, N.L./Beswick, R.W., 2001: Intersubjective Semantic Meanings Emergent in a Work Group: A Neural Network Content Analysis of Voice Mail. S.33-50 in: M. D. West: Applications of Computer Content Analysis. Norwood, NJ: Ablex.

SPSS Inc., 1997: TextSmart User's Guide. Chicago: Spss Inc.

Stone, P.J./Dunphy, D. C./Smith, M. S./Ogilvie, D. M., 1966: The General Inquirer: A Computer Approach to Content Analysis. Cambridge, Massachusetts, London: The M. I. T. Press.

Woelfel, J.K., 1998: User's Guide. Catpac II. New York: Rah Press.

Züll, C./Landmann, J., 2002: Computerunterstützte Inhaltsanalyse: Literaturbericht zu neueren Anwendungen. Mannheim: ZUMA.

http://www.gesis.org/Publikationen/Berichte/ZUMA_Methodenberichte/documents/pdfs/tb02_02.pdf.

Korrespondenzadressen

Juliane Landmann
ZUMA
Postfach 12 21 55
D-68072 Mannheim
email: landmann@zuma-mannheim.de

Cornelia Züll
ZUMA
Postfach 12 21 55
D-68072 Mannheim
email: zuell@zuma-mannheim.de

ERFAHRUNGEN MIT PERSONENNAMEN ZUR BILDUNG VON STICHPROBEN FÜR BETRIEBSBEFRAGUNGEN

ANDREAS HUMPERT

Im vorliegenden Beitrag werden in kurzer Form die Vorgehensweise und die Erfahrungen beim Einsatz von Personennamen zur Bearbeitung des öffentlichen Telefonverzeichnisses vorgestellt. Im Forschungsprojekt "Umfang, Entwicklung und Potentiale an Einfacharbeitsplätzen in der Region Rhein-Neckar" musste eine Auswahlgrundlage für die telefonische Befragung von Betrieben gebildet werden. Aufgrund der Unvollständigkeit oder eingeschränkten Verfügbarkeit vieler Betriebs- und Unternehmensverzeichnisse wurde deshalb auf das öffentliche Telefonverzeichnis zurückgegriffen. Das zentrale Argument für die Nutzung des Telefonbuches bestand in der Annahme, dass fast alle Betriebe in dieser bundesweiten Datengrundlage eingetragen sind. Für die Ziehung einer Betriebsstichprobe musste jedoch noch eine weitere Bearbeitung erfolgen, die eine gleiche Auswahlwahrscheinlichkeit gewährleisten und neutrale Ausfälle bzw. Fehlkontakte in der Feldphase minimieren sollte. Unter Berücksichtigung der eingeschränkten Bedingungen für die Stichprobenziehung im Rahmen des genannten Erhebungsprojektes sind mit der Vorgehensweise erste positive Erfahrungen gesammelt worden, die zu einer weiteren Prüfung veranlassen sollen.

The article presents a brief account of our experiences in using people's last names as found in public telephone directories for a telephone survey of businesses. In the research project "Range, Development and Potentials of Simple Job Positions in the Rhein-Neckar Region", a new set of rules for selection of businesses had to be made for telephone interviews. Many directories specifically for business and firm numbers are incomplete or have limited access, making it necessary to resort to the public telephone directory to locate more firms and businesses. The key reason for using the general telephone directory is the assumption that almost all businesses are registered in this nationwide data base. In order to draw a sample of businesses, the data in the public telephone directory had to be edited so as to guarantee equal probability of selection and minimize

ineligibles and wrong numbers in the field phase. Given the restrictions to the sampling design dictated by the nature of the project, these first findings for using the public telephone directory were positive enough to warrant further research into its potential.

1. Einleitung

Im vorliegenden Text¹ werden in kurzer Form die Vorgehensweise und die Erfahrungen beim Einsatz von Personennamen zur Bearbeitung des öffentlichen Telefonverzeichnisses vorgestellt. Für die telefonische Befragung im Rahmen des Projektes "Umfang, Entwicklung und Potentiale an Einfacharbeitsplätzen in der Region Rhein-Neckar"² musste eine Auswahlgrundlage für eine Betriebsstichprobe gebildet werden. Es wurde auf das öffentliche Telefonverzeichnis zurückgegriffen. Für die Ziehung einer Stichprobe mit Betrieben musste jedoch noch eine weitere Bearbeitung erfolgen. In den folgenden Erläuterungen werden die notwendigen Bearbeitungsschritte und erste Erfahrungen dargestellt.

2. Projektspezifische Rahmenbedingungen für die Auswahlgrundlage

Die hier dargestellte Vorgehensweise bei der Bildung einer Auswahlgrundlage für eine Befragung von Betrieben mit Beschäftigten muss vor den einschränkenden Rahmenbedingungen universitärer Forschung beurteilt werden. Im Rahmen des einjährigen Forschungsprojektes "Umfang, Entwicklung und Potentiale an Einfacharbeitsplätzen in der Region Rhein-Neckar" am Institut für Mittelstandsforschung (ifm) der Universität Mannheim sollten 1.000 Betriebe aus allen Wirtschaftszweigen befragt³ werden. Die Grundgesamtheit war durch alle Betriebe mit Beschäftigten (einschließlich nicht-sozialversicherungspflichtig Beschäftigter) im Untersuchungsgebiet definiert. Das Untersuchungsgebiet war durch die Grenzen des Raumordnungsverband Rhein-Neckar (ROV)⁴ festgelegt. Der Begriff des "Betriebes" ist von dem Unternehmen als wirtschaftlich-rechtliches Gebilde zu unterscheiden. Unter "Betrieb" wird die örtliche Einheit verstanden, in denen die Tätigkeiten eines Unternehmens, d.h. die Produktion von Gütern oder

1 Ich danke Herrn Prof. Dr. Borg und anonymen Gutachtern dieser Zeitschrift für die nützlichen Korrekturhinweise.

2 Das Projekt wurde vom Institut für Mittelstandsforschung (ifm) der Universität Mannheim durchgeführt und von der EQUAL-Entwicklungspartnerschaft Rhein-Neckar-Dreieck, AG 2 finanziell gefördert.

3 Die Erhebung wurde von ZUMA (Pretest) und dem Sozialwissenschaftlichen Umfragezentrum der Universität Duisburg-Essen zwischen dem 11.11. und 18.12.03 durchgeführt.

4 Nähere Information unter: <http://www.region-rhein-neckar-dreieck.de/fakten/region.html>

Dienstleistungen tatsächlich durchgeführt werden. Personal- und Beschäftigungsentscheidungen werden meist auf dieser Ebene ausgeführt.

Die Wahl einer Auswahlgrundlage für die Stichprobenziehung war aber nicht nur durch diese zeitlichen, finanziellen und thematischen Bedingungen eingeschränkt. Auch methodisch wurde mit der Wahl der telefonisch-mündlichen Erhebungsart (CATI) eine weitere Bedingung gesetzt. Die stärkeren Kontrollmöglichkeiten, die kürzere Feldphase, flexiblere Feldsteuerung und die höhere Teilnahmereitschaft der Zielgruppe bei einem telefonischen Interview sprachen neben anderen Argumenten (vgl. Häder 2000: 2) für diese Entscheidung. Um eine Auswahlgrundlage zu bilden, welche mit der Struktur der eigentlichen Grundgesamtheit der beschriebenen Zielgruppe möglichst übereinstimmte, sollten deshalb im genannten Befragungsprojekt folgende Kriterien bei der Wahl der Datenquelle gelten:

- die Grundgesamtheit sollte besonders hinsichtlich der Untersuchungsregion und der Wirtschaftszweigstruktur möglichst vollständig enthalten sein
- die Kontaktdaten der Betriebe sollten für eine universitäre Forschungseinrichtung verfügbar sein
- die Kosten für die Stichprobenziehung sollten im Projektrahmen liegen
- das Auswahlverfahren sollte nachvollziehbar bzw. kontrollierbar sein
- der Telefonanschluss sollte in der Datenquelle enthalten sein
- jedes Element der Grundgesamtheit sollte die gleiche Auswahlchance besitzen
- es sollten aktuelle bzw. "gepflegte" Daten sein (vgl. Brüderl 1996:67ff)

Vor dem Hintergrund dieser projektspezifischen Bedingungen für die Entscheidung über die Auswahlgrundlage waren viele Firmen- bzw. Betriebsverzeichnisse nicht geeignet. Umfangreiche und nahezu vollständige Betriebsverzeichnisse wie die BA-Betriebsdatei oder auch das Unternehmensregister des Statistischen Bundesamtes blieben wegen fehlender Verfügbarkeit sowieso unberücksichtigt. Bei Verzeichnissen der Kammern (IHK, HK, berufsständische Kammern) war mit einer unvollständigen Besetzung bestimmter Wirtschaftszweige, höheren Kosten, fehlenden Rufnummern oder auch einem älteren Datenstand zu rechnen. Auch auf kommerzielle Anbieter von Firmendaten wie z.B. Hoppenstedt, Schober oder Creditreform wurde wegen der zu erwartenden systematischen Über- bzw. Untererfassung bestimmter Firmentypen nicht zurückgegriffen. Aber auch die fehlende Nachvollziehbarkeit und Dokumentation bei der Zusammenspielung (vgl. Salentin 1994: 65) verschiedenster Datenquellen (z.B. Handelsregister, Inkassodaten, etc.) war neben den höheren Kosten ein schwerwiegendes Argument dafür, dass Auswahlgrundlagen kommerzieller Adressanbieter nicht für die Betriebsstichprobe genutzt wurden. Unter den gegebenen Rahmenbedingungen des "Einfacharbeitsplatz"-Projektes erfüllte deshalb das öffentliche Telefonverzeichnis (Ausgabe 2003) als Datenquelle die

meisten Kriterien für die erforderliche Bildung einer geeigneten Auswahlgrundlage. Hinsichtlich dieser Entscheidung über die Wahl der Datenquelle sind jedoch drei Aspekte zu berücksichtigen:

1. Die Grundgesamtheit ist (fast) vollständig in der Datenquelle (Telefonbuch) enthalten:

Dies ist lediglich eine theoretische Annahme, die mit dem Interesse der Unternehmen bzw. Betriebe nach leichter Erreichbarkeit (z.B. durch Kunden) begründet werden kann. Vor diesem Hintergrund ist der Eintrag eines Unternehmens im öffentlichen Telefonverzeichnis im Allgemeinen anzunehmen. Allerdings kann dies nicht für alle Betriebe unterstellt werden. Beispielsweise wären folgende Motive für die Nichteintragung eines Betriebes denkbar:

- verschiedene Formen illegaler gewerblicher bzw. selbständiger Tätigkeit
- es werden keine (weiteren) Kunden gesucht bzw. Kundenkontakt ist nicht erwünscht
- der Betrieb ist Teil eines größeren Unternehmens mit vorhandenem Telefonbucheintrag

Zwar kann angenommen werden, dass nur eine Minderheit der Grundgesamtheit nicht über das öffentliche Telefonverzeichnis erreichbar ist. Dennoch wäre hinsichtlich dieses Problempunktes zukünftig zu überprüfen, ob sich diese Minderheit systematisch von der Grundgesamtheit unterscheidet.

2. Eintragung von Nicht-Betrieben:

Im Telefonbuch sind neben der Zielgruppe auch Elemente eingetragen, die nicht zur Grundgesamtheit gehören. Dies sind mehrheitlich Privathaushalte und auch Nicht-Privathaushalte, die aber nicht als Betrieb (z.B. Sportplatz, Gemeindesaal etc.) angesehen werden können. Diese Telefonbucheinträge verursachen in der Feldphase zeitlichen und finanziellen Mehraufwand wegen der daraus resultierenden Fehlkontakte, die bei mindestens $\frac{3}{4}$ einer Bruttostichprobe zu erwarten wären. Ein Ausschluss von Einträgen, die keine Betriebe sind, wäre also aus Sicht der Erhebungskosten erforderlich.

3. Mehrfacheintragung von einzelnen Betrieben:

Viele Betriebe stehen mit mehreren Anschlüssen (Durchwahlen, Faxanschlüsse, verschiedenen Betriebsabteilungen etc.) im öffentlichen Telefonverzeichnis, und erhöhen damit natürlich ihre Auswahlchance. Eine systematische Überrepräsentierung z.B. größerer Betriebe mit mehreren (eingetragenen) Anschlüssen wäre zu erwarten. Deshalb ist eine vorbereitende Bearbeitung des Telefonbuches, indem entsprechende Anschlusstypen oder mehrfache Betriebseintragungen unterbunden werden, unerlässlich.

Der erste Aspekt wurde im "Einfacharbeitsplatz"-Projekt als zentrale Annahme eingeführt. Eine empirische Überprüfung konnte unter den Rahmenbedingungen des Projektes nicht geleistet werden, sondern wäre eine Aufgabenstellung nachfolgender Methodentests. Bei der vorbereitenden Bearbeitung des öffentlichen Telefonverzeichnisses mussten jedoch die angesprochenen Probleme in Punkt 2 und 3 weitestgehend gelöst werden. Mit den folgenden Ausführungen soll die Aufmerksamkeit auf die spezifische Vorgehensweise bei der Datenbearbeitung zur Lösung dieser beiden Problempunkte gelenkt werden.

3. Vorbereitung des Telefonverzeichnisses als Auswahlgrundlage

Für die Lösung der Problempunkte 2 und 3 sind folgende zwei Bearbeitungsschritte notwendig:

1. Bearbeitungsschritt:

Das Löschen von Einträgen, die offenkundig nicht zur Zielgruppe (Betriebe mit Beschäftigten) gehören, konnte mit Hilfe von Verzeichnissen aus der Namenforschung (Onomastik) erreicht werden (vgl. Humpert/Schneiderheinze 2000,2002)⁵. Grundsätzlich wurde davon ausgegangen, dass die ausschließliche Eintragung von Personennamen (Vor- und Nachnamen) auf den Telefonanschluss eines Privathaushaltes hindeutet. Auf der Grundlage von 1,1 Mio. Personennamen⁶, die aus Monographien, Wörterbüchern und Lexika der Namenforschung stammen⁷, wurden alle Anschlüsse mit ausschließlicher Eintragung von Personennamen gelöscht. Dabei war allerdings zu beachten, dass eine sehr kleine Zahl von Nachnamen mit der Bezeichnung eines Betriebes übereinstimmen (z.B. "Hotel"). Deshalb wurde bei solchen Übereinstimmungen zwischen Nachnamen und Betriebs- bzw. Branchenbezeichnungen der entsprechende Name nicht berücksichtigt. Das versehentliche Löschen von Betrieben, deren Branchenbezeichnung im Telefonbuch fehlte⁸, konnte durch den Abgleich verschiedener digitaler Telefonverzeichnisse weitestgehend vermieden werden.

5 Im Zusammenhang mit der Zuordnung von Personennamen zur sprachlichen Herkunft.

6 Bestehend aus rund 900.000 Nachnamen und rund 200.000 Vornamen aus insgesamt 89 Sprachen.

7 Aus über 200 Veröffentlichungen der Namenforschung wurden alle Vor- und Nachnamen nach sprachlicher Herkunft in einer Datenbank mit mittlerweile 1,1 Mio. Eintragungen eingegeben. Eine Übersicht der genutzten Literatur ist als PDF-Dokument unter <http://www.stichproben.de> erhältlich.

8 Beispielsweise durch die unterschiedliche Struktur verschiedener digitaler Telefonverzeichnisse, die nach Anlage eines Branchenverzeichnisses den entsprechenden Begriff löschen.

2. Bearbeitungsschritt:

Die verbliebenen Eintragungen des öffentlichen Telefonverzeichnisses wurden hinsichtlich mehrfacher Eintragungen einzelner Betriebe bereinigt. Dabei wurden mehrheitlich Fax- und Mobilfunkanschlüsse gelöscht, wenn der Betrieb noch mit einem Festnetzanschluss eingetragen war. Dazu war ein Abgleich aller Eintragungen nach unterschiedlichen Sortierungen (Adresse, ähnliche Namenseintragungen etc.) erforderlich. Zusätzlich wurde auch die mehrfache Eintragung (oftmals mit derselben Rufnummer) und aufgrund von verzeichneten Durchwahlen (zu verschiedenen Abteilungen eines Betriebes) hervorgerufene Mehrfacheintragungen beseitigt. Mehrere Betriebe des selben Unternehmens bzw. Unternehmensgruppe wurden nicht gelöscht, da die für die Untersuchungsthematik relevanten Entscheidungen (z.B. Personalentscheidungen) auf betrieblicher Ebene getroffen werden.

Im Projekt "Umfang, Entwicklung und Potentiale an Einfacharbeitsplätzen in der Region Rhein-Neckar" führten die beiden Bearbeitungsvorgänge zu den folgenden Reduzierungen des kompletten Bestandes des öffentlichen Telefonverzeichnisses⁹ in der Untersuchungsregion (ROV)¹⁰.

Tabelle 1: Vorbereitung des Telefonverzeichnisses

	Datensätze
Untersuchungsgebiet (ROV)	906.136
<i>Löschung folgender Einträge:</i>	
nur Personennamen enthalten	-700.671
2./3.usw. Eintrag eines Betriebes/Haushaltes (Faxe, Mobil, Durchwahlen etc.)	-51.267
verbleibende Auswahlgrundlage	154.198

Aus den verbliebenen 154.198 Datensätzen wurde eine reine Zufallsauswahl von **7.000** Rufnummern getroffen. Auf eine Schichtung nach regionaler Verteilung wurde verzichtet, um die Möglichkeit einer nachträglichen Kontrolle durch einen Vergleich mit den amtlichen Daten (BA) zu erhalten. Dadurch war es möglich, dass jeweils die Abweichungen der Bruttostichprobe und der Befragtengruppe zur Grundgesamtheit erfasst werden konnten. Zumindest hinsichtlich der regionalen Verteilung war damit überprüfbar, ob Abwei-

⁹ Stand: Januar 2003

¹⁰ Der Raumordnungsverband Rhein-Neckar besteht aus 4 Landkreisen und 7 kreisfreien Städten.

chungen zur Grundgesamtheit durch das Stichprobenverfahren oder die Erhebung (Ausschöpfung) verursacht wurden.

Da die Listen mit den Personennamen nicht vollständig sind, musste diese Bruttostichprobe in einer Einzelfallprüfung nochmals bereinigt werden. Dabei wurden solche Eintragungen gelöscht, die nur einen Vor- und Nachnamen enthielten. In Einzelfällen wurden auch Datensätze mit Berufseinträgen gelöscht, die eindeutig auf einen Privathaushalt (z.B. "Studienrat a.D.") hindeuteten. Letztlich blieben danach noch **5.414** Rufnummern übrig, von denen **5.204** zufällig ausgewählt für die Erhebung als Bruttostichprobe eingesetzt wurden.

4. Erste Erfahrungen mit der Betriebsstichprobe

Mit dem praktischen Einsatz dieser Bruttostichprobe im "Einfacharbeitsplatz"-Projekt konnten auch erste Erfahrungen hinsichtlich der Effizienz und Fehlerhaftigkeit des Vorgehens gewonnen werden. Als grobe Indikatoren beider Aspekte können das Ausschöpfungsprotokoll und ein Vergleich der Befragtengruppe mit amtlichen Daten zur Grundgesamtheit gewertet werden, - obwohl dies natürlich einen notwendigen Methodentest nicht ersetzen kann.

Die Löschung von Elementen, welche nicht zur Zielgruppe gehören (vorwiegend Privathaushalte und Betriebe ohne Beschäftigte), ist möglicherweise nicht hinreichend durch das Kennzeichnen mit Personennamen aus Verzeichnissen der Namenforschung (Onomastik) zu lösen. Die Unvollständigkeit der Namenlisten dürfte ein erster Grund für die unvollständige Säuberung der Auswahlgrundlage sein. Dies konnte man bereits daran erkennen, dass unter den 7.000 Rufnummern aus der bereinigten Auswahlgrundlage durch die Einzelfallprüfung noch immerhin **1.586** Datensätze mit ausschließlich Personennamen kenntlich gemacht werden konnten. Eine Erweiterung der Namenlisten (aktuell: 1,1 Millionen Vor- und Nachnamen) wäre deshalb sicherlich notwendig. Allerdings konnte ein größerer Teil von den 1.586 in der Einzelfallprüfung gelöschten Einträgen durch Namenlisten nicht gefunden werden, weil auch Berufs- und Bildungsangaben von eindeutigen Privatpersonen (z.B. "Direktor a.D.") im Telefonbuch enthalten sind. Das folgende Ausschöpfungsprotokoll verdeutlicht jedoch auch, dass trotz umfassender Namenslisten **und** Einzelfallprüfung ein nennenswerter Anteil (17,4%) der Bruttostichprobe noch aus Privathaushalten bestand. Dieser Mehraufwand bei der Kontaktaufnahme war im Rahmen dieses projektgebunden Vorgehens (noch) nicht zu vermeiden. Die Analyse der Einträge dieser 905 Rufnummern von Privathaushalten ergab jedoch, dass sich darunter mehrheitlich Privatpersonen mit Berufs- und Bildungseintragungen befanden. Daraus ergeben sich Ansatzpunkte für eine zukünftige Verbesserung der Vorgehensweise.

Tabelle 2: Ausschöpfung der Bruttostichprobe

	absolut	relativ (in %)
<i>Rufnummern gesamt</i>	5.204	100,0
kein Anschluss	337	6,5
Nummer geändert	10	0,2
Fax, Modem	154	3,0
kein Betrieb	905	17,4
Betrieb hat keine Beschäftigten	933	17,9
Summe	2.339	45,0
<i>Bereinigte Bruttostichprobe</i>	2.865	100,0
Freizeichen/kein Kontakt (nach 10 Versuchen)	142	5,0
Anrufbeantworter	101	3,5
ständig besetzt	14	0,5
verschiedene Verweigerungsformen	1.531	53,4
Verständigungsschwierigkeiten	73	2,6
<i>Interviews</i>	1004	35,0

Die hohen Ausfälle durch "fehlende Beschäftigte" sind jedoch auch zukünftig kaum zu vermeiden, da die entsprechenden Datensätze im Telefonverzeichnis nur schwer identifizierbar sind. Diese Einschränkung gilt allerdings nicht, wenn auch Betriebe ohne Beschäftigte (Soloselbständige) zur Grundgesamtheit zählen.

Der Vergleich der befragten Betriebe mit amtlichen Daten zur Grundgesamtheit (vgl. Tabellen 3-6) gibt erste Hinweise darauf, ob durch das Verfahren eine Auswahlgrundlage gebildet wird, die systematisch von der Grundgesamtheit abweicht. Die daraus gezogenen Stichproben würden ebenfalls systematisch von der Struktur der Grundgesamtheit abweichen. Allerdings kann aufgrund des hohen Anteils von Ausfällen nicht ausgeschlossen werden, dass die Zusammensetzung der Befragtengruppe auch durch Verweigerung bzw. Nichtteilnahme systematisch verändert wurde.

Für den Vergleich konnten nur 907 der 1.004 befragten Betriebe herangezogen werden, da sich die amtlichen Rahmendaten der BA (Betriebsstatistik 2001) nur auf die Betriebe mit mindestens einem sozialversicherungspflichtigen Beschäftigten beziehen. In der Befragung wurden jedoch auch solche Betriebe erfasst, die Beschäftigte ohne Sozialversicherungspflicht haben. Außerdem ist zu berücksichtigen, dass sich die Daten zur Grundge-

samtheit und den Primärdaten auf unterschiedliche Zeitpunkte (mehr als zwei Jahre Unterschied) beziehen. Da aus der Betriebsdatei der Beschäftigtenstatistik nur die Verteilung nach Region, Beschäftigtenzahl und Wirtschaftszweig verfügbar war, beschränkt sich der Vergleich auch nur auf diese drei Merkmale. Es soll auch nicht unerwähnt bleiben, dass selbst bei der Beschäftigtenstatistik der Bundesanstalt für Arbeit (BA), die auf einer Pflichtmeldung der Arbeitgeber an die Sozialversicherungsträger beruht, hinsichtlich kleinbetrieblicher Strukturen deutliche Unschärfen zu beobachten sind (vgl. Fritsch 2002:213).

Hinsichtlich der regionalen Verteilung war zusätzlich auch noch ein Vergleich der regionalen Verteilung in der Bruttostichprobe (5.204 Rufnummern) und der Befragtengruppen (907 Befragte) möglich, um mögliche Ursachen für Abweichungen zur Grundgesamtheit offenzulegen.

Tabelle 3: Vergleich der regionalen Verteilung der Bruttostichprobe und Grundgesamtheit

Kreis	KSN	Bruttostichprobe***	Grundgesamtheit**	Differenz (in Prozentpunkten)
Bad Dürkheim LK	07 332	7,6%	7,0%	+ 0,6
Bergstrasse LK	06 431	15,2%	14,3%	+ 0,9
Ludwigshafen LK	07 338	6,1%	6,2%	- 0,1
Rhein-Neckar-Kreis LK	08 226	27,4%	26,2%	+ 1,2
Frankenthal Kf	07 311	2,2%	2,3%	- 0,1
Heidelberg Kf	08 221	9,2%	8,9%	+ 0,3
Ludwigshafen Kf	07 314	6,4%	6,9%	- 0,5
Mannheim Kf	08 222	15,5%	17,1%	- 1,6
Neustadt Kf	07 316	3,6%	3,5%	+ 0,1
Speyer Kf	07 318	2,9%	3,1%	- 0,2
Worms Kf	07 319	4,0%	4,4%	- 0,4

Tabelle 4: Vergleich der regionalen Verteilung bei befragten Betrieben und Grundgesamtheit

Kreis	KSN	befragte Betriebe*	Grundgesamtheit**	Differenz (in Prozentpunkten)
Bad Dürkheim LK	07 332	7,9%	7,0%	+ 0,9
Bergstrasse LK	06 431	13,2%	14,3%	- 1,1
Ludwigshafen LK	07 338	5,6%	6,2%	- 0,6
Rhein-Neckar-Kreis LK	08 226	28,9%	26,2%	+ 2,7
Frankenthal Kf	07 311	1,7%	2,3%	- 0,6
Heidelberg Kf	08 221	10,3%	8,9%	+ 1,4
Ludwigshafen Kf	07 314	5,3%	6,9%	- 1,6
Mannheim Kf	08 222	15,3%	17,1%	- 1,8
Neustadt Kf	07 316	4,6%	3,5%	+ 1,1
Speyer Kf	07 318	4,1%	3,1%	+ 1,0
Worms Kf	07 319	3,1%	4,4%	- 1,3

LK = Landkreis Kf = kreisfreie Stadt // KSN = Kreisschlüsselnummer des Statistischen Bundesamtes

* Basis: 907 befragte Betriebe mit sozialversicherungspflichtig Beschäftigten (Erhebung im Projekt "Einfacharbeitsplätze im ROV" 2003)

** Basis: 48.081 Betriebe mit sozialversicherungspflichtig Beschäftigten im ROV
Quelle: Beschäftigtenstatistik 2001, BA-Betriebsdaten

*** Basis: 5.204 zufällig ausgewählte Betriebe (Bruttostichprobe)

Beim Vergleich von Bruttostichprobe und Befragtengruppe (Tabelle 3 und 4) wird ersichtlich, dass die regionale Verteilung der Grundgesamtheit in der Bruttostichprobe deutlich besser abgebildet wird. Die vorhandenen (geringen) Abweichungen in der Bruttostichprobe wurden also zusätzlich im Rahmen des Erhebungsprozesses (Ausschöpfung der Bruttostichprobe) noch vergrößert. Dies kann als Indiz dafür gewertet werden, dass die verzerrte Abbildung der Struktur in der Grundgesamtheit auch durch erhebungsbedingte Faktoren (z.B. Befragungsart, Erhebungszeitraum, Erreichbarkeit der Zielgruppe, Interviewer etc.) verursacht wird.

Tabelle 5: Vergleich der Beschäftigtenzahl bei befragten Betrieben und Grundgesamtheit

Zahl der sozialversicherten Beschäftigten	befragte Betriebe*	Grundgesamtheit**	Differenz (in Prozentpunkten)
1 - 4	46,9%	64,3%	- 17,4
5 - 9	21,3%	17,1%	+ 4,2
10 - 19	12,5%	9,0%	+ 3,5
20 - 49	9,7%	5,7%	+ 4,0
50 - 99	4,6%	2,1%	+ 2,5
100 - 249	2,4%	1,2%	+ 1,2
250 und mehr	2,6%	0,7%	+ 1,9

Tabelle 6: Vergleich der Beschäftigtenzahl bei befragten Betrieben und Grundgesamtheit

Wirtschaftszweig	befragte Betriebe*	Grundgesamtheit**	Differenz (Prozentpunkte)
Land- und Forstwirtschaft	3,1%	2,3%	+ 0,8
Bergbau	0,1%	0,2%	- 0,1
Verarbeitendes Gewerbe	12,2%	11,0%	+ 1,2
Energie- und Wasserversorgung	0,2%	0,2%	± 0,0
Baugewerbe	8,0%	10,4%	- 2,4
Handel	21,3%	22,7%	- 1,4
Gastgewerbe	5,5%	7,4%	- 1,9
Verkehr- und Nachrichtenübermittlung	4,6%	5,6%	- 1,0
Kredit- und Versicherungsgewerbe	3,6%	2,3%	+ 1,3
Dienstleistungen für Unternehmen	18,1%	16,9%	+ 1,2
Öffentliche Verwaltung, Sozialversicherung	2,2%	1,3%	+ 0,9
Erziehung und Unterricht	3,2%	2,4%	+ 0,8
Gesundheits-, Veterinär- und Sozialwesen	10,6%	10,5%	+ 0,1
Erbringung sonstiger Dienstleistungen	7,1%	6,9%	+ 0,2

LK = Landkreis Kf = kreisfreie Stadt // KSN = Kreisschlüsselnummer des Statistischen Bundesamtes

* Basis: 907 befragte Betriebe mit sozialversicherungspflichtig Beschäftigten (Erhebung im Projekt "Einfacharbeitsplätze im ROV" 2003)

** Basis: 48.081 Betriebe mit sozialversicherungspflichtig Beschäftigten im ROV

Quelle: Beschäftigtenstatistik 2001, BA-Betriebsdaten

Zusammenfassend kann anhand der Vergleichsdaten festgehalten werden, dass sich die regionale Verteilung und Wirtschaftszweigstruktur der befragten Betriebe geringfügig von den Verteilungen in der Grundgesamtheit unterscheiden. Dies kann für die Betriebsgröße **nicht** beobachtet werden. Die Gruppe der kleinsten Betriebe (1 bis 4 Beschäftigte) ist in den Primärdaten deutlich unterrepräsentiert. Allerdings kann hier nicht gesagt werden, ob nicht auch der höhere Zeitdruck und die schlechtere Erreichbarkeit der Entscheidungsträger in Kleinstbetrieben zu höheren Ausfällen in dieser Betriebsgrößengruppe geführt hat. Zumindest beim Merkmal "regionale Verteilung" (Kreisebene) konnte festgestellt werden, dass auch durch erhebungsbedingte Faktoren eine korrekte Abbildung der Verteilungen in der Grundgesamtheit beeinträchtigt wird. Andererseits ist auch nicht auszuschließen, dass verstärkt kleine Betriebe nicht im Telefonbuch stehen oder als Betrieb nicht erkennbar sind. Abschließend ist zu berücksichtigen, dass bei den Vergleichen mit höherer Zahl von Merkmalsausprägungen auch geringere Abweichungen beobachtet wurden.

5. Schlussfolgerungen

Das dargestellte Vorgehen zur Bildung einer Auswahlgrundlage für die telefonische Befragung von Betrieben resultierte vorwiegend aus den Rahmenbedingungen des Forschungsprojektes. Üblicherweise werden Auswahlgrundlagen für Betriebsumfragen durch die gezielte Erfassung und Zusammentragung der Unternehmen gebildet. Dabei werden oft ausgewählte Branchen bzw. Wirtschaftszweige zur Zielgruppe erklärt, für die auch umfassende Adressverzeichnisse vorliegen. Häufig wird auch ein Design aus Gebietsstichprobe und zufälliger Auswahl der Betriebe aus Kammerverzeichnissen gewählt. Entsprechend häufig waren viele Untersuchungskonzepte schon darauf ausgelegt, dass bestimmte Wirtschaftszweige nicht zur Zielgruppe gehören, da keine geeigneten Verzeichnisse bzw. Auswahlgrundlagen vorlagen. Im vorgestellten Projekt war man hingegen darum bemüht, dass alle Wirtschaftszweige in die Stichprobe gelangen konnten. Deshalb bezogen sich die Arbeiten nicht auf das Sammeln von Zielgruppenelementen zur Bildung einer Auswahlgrundlage, sondern auf die Löschung von Elementen, die nicht zur Zielgruppe gehörten, aus einer (fast) vollständigen Auswahlgrundlage.

Hier sollten zunächst nur erste Erfahrungswerte geschildert werden, um eine Anregung für andere Untersuchungen und kleine Forschungsgruppen mit ähnlicher Aufgabenstellung für die Stichprobenziehung zu bieten. Es ist deshalb unmittelbar einsichtig, dass hier keine perfekte Lösung geboten werden konnte. Vielmehr sollte dieser Projektbericht auch auf mögliche Schwächen der Vorgehensweise (z.B. Nichteintrag von Betrieben im öffentlichen Telefonverzeichnis; Unterrepräsentierung kleinster Betriebe) hinweisen, um zu weiteren Kontrollen anzuregen.

Literatur

Brüderl, J./Preisendörfer, P. Ziegler, R., 1996: Der Erfolg neugegründeter Betriebe. Berlin: Duncker & Humblot.

Fritsch, M./Grotz, R./Brixy, U./Niese, M./Otto, A., 2002: Zusammenfassender Vergleich der Datenquellen zum Gründungsgeschehen in Deutschland, S. 199-214, in: M. Fritsch/R. Grotz (Hrsg.), Das Gründungsgeschehen in Deutschland. Heidelberg: Physica-Verlag.

Häder, S., 2000: Telefonstichproben. ZUMA How-to-Reihe 6.

Humpert, A./Schneiderheinze, K., 2000: Stichprobenziehung für telefonische Zuwandererumfragen - Einsatzmöglichkeiten der Namenforschung. ZUMA-Nachrichten 47: 36-63.

Humpert, A./Schneiderheinze, K., 2002: Stichprobenziehung für telefonische Zuwandererumfragen - Praktische Erfahrungen und Erweiterungen der Auswahlgrundlage, S. 187-208, in: Gabler, S./Häder, S. (Hrsg.), Telefonstichproben - methodische Innovationen und Anwendung in Deutschland. Münster: Waxmann.

Salentin, K., 1994: Direktmarketingadressen für Umfragezwecke?. ZUMA-Nachrichten 35: 54-65.

Korrespondenzadresse

*Dr. Andreas Humpert
Institut für Mittelstandsforschung (ifm)
der Universität Mannheim
D 6,5
D-68159 Mannheim
email: humpert@mail.ifm.uni-mannheim.de
<http://www.ifm.uni-mannheim.de>*

ZUR UMSETZUNG DES INTERNATIONALEN SOZIOÖKONOMISCHEN INDEX DES BERUFLICHEN STATUS (ISEI) MIT DEN MIKROZENSEN AB 1996

BERNHARD SCHIMPL-NEIMANNS

In ländervergleichenden Analysen wird als Indikator der sozioökonomischen Lage häufig der Internationale Sozioökonomische Index des beruflichen Status (International Socio-Economic Index of Occupational Status; ISEI) verwendet. Die Konstruktion von ISEI setzt Berufsangaben voraus, die nach der Internationalen Standardklassifikation der Berufe 1988 (ISCO-88) verschlüsselt sind. In diesem Beitrag wird gezeigt, wie ISEI für die Mikrozensus-Scientific Use Files umgesetzt werden kann, in denen diese Informationen ab 1996 vorliegen. Darüber hinaus werden Zusammenhänge zwischen ISEI und weiteren sozioökonomischen Variablen untersucht. Hierbei zeigt sich, dass bei statistischer Kontrolle dieser anderen Variablen der Zusammenhang zwischen ISEI und dem Einkommen sehr gering ist. Die Ergebnisse können dazu beitragen, die Möglichkeiten und Grenzen der Verwendung von ISEI besser zu beurteilen.

The International Socio-Economic Index of Occupational Status (ISEI) is frequently used as an indicator of socio-economic position in comparative analyses. The construction of ISEI status requires that occupations are coded according to the International Standard Classification of Occupations (ISCO-88). This occupational information has been part of the German Mikrozensus Scientific Use Files since 1996. This paper describes how ISEI can be applied to the German Mikrozensus. Associations between ISEI and a number of other socio-economic variables are also analyzed. These demonstrate that the partial correlation between ISEI and income is rather small. The findings can be seen as a contribution to assessing the potential and limitations of ISEI in analysis.

1. Einleitung¹

Für die Sozial- und Wirtschaftswissenschaften sind Angaben zur Erwerbstätigkeit im Allgemeinen und zum ausgeübten Beruf im Besonderen von großer Bedeutung, da die Stellung auf dem Arbeitsmarkt in enger Verbindung mit der materiellen und sozialen Lage sowie den Lebenschancen der Bevölkerung steht. Die Scientific Use Files des Mikrozensus enthalten die nach der deutschen Klassifizierung der Berufe codierten Berufsangaben. Ab dem Erhebungsjahr 1996 liegen zusätzlich die nach der Internationalen Standardklassifikation der Berufe in der Fassung zur Verwendung innerhalb der Europäischen Gemeinschaft (ISCO-88 COM) vercodeten Angaben vor. Dies hat die Verwendung des Mikrozensus für ländervergleichende Analysen entscheidend verbessert. Darüber hinaus ist es damit möglich, in der international vergleichenden Sozialforschung verwendete Klassifikationen und Skalen zu konstruieren, die auf ISCO-88 basieren. Dazu zählt insbesondere der von Ganzeboom und Treiman entwickelte Internationale Sozioökonomische Index des beruflichen Status (Socio-Economic Index of Occupational Status; ISEI) (vgl. Ganzeboom et al. 1992; Ganzeboom/Treiman 1996; Ganzeboom/Treiman 2003). Unter anderem wurde ISEI in PISA bei den internationalen Vergleichen als Standardindikator für die sozioökonomische Stellung der Eltern (Baumert et al. 2001: 327) sowie in Analysen der Europäischen Arbeitskräftestichproben eingesetzt (Handl/Steinmetz 2003; Kogan/Müller 2003; Müller/Gangl 2003).

Bei der Optimierung von Skalen und Klassifikationen für international vergleichende Analysen können nationale Besonderheiten nur ansatzweise berücksichtigt werden. Für die Anwendung von ISEI mit Mikrozensusdaten ist es deshalb hilfreich, etwas über die Passfähigkeit von ISEI zu erfahren.

Im nächsten Abschnitt werden zunächst die zentralen Eigenschaften von ISEI skizziert. Daran anschließend folgt im dritten Abschnitt eine kurze Beschreibung, wie ISEI im Mikrozensus gebildet werden kann. Im vierten Abschnitt werden mit dem Ziel einer ersten Validierung auf Basis des Mikrozensus 2000 Zusammenhänge zwischen ISEI und weiteren sozioökonomischen Variablen untersucht. Die für die Umsetzung von ISEI notwendigen Umsteigeschlüssel von ISCO-88 (COM) zu ISEI sowie ein SPSS-Programm mit Recodierungsanweisungen sind als Mikrodaten-Tools im WWW zu finden. (URL: http://www.gesis.org/Dauerbeobachtung/Mikrodaten/mikrodaten_tools/ISEI/isei_tools.htm).

1 Für hilfreiche Anregungen zu einer früheren Fassung danke ich Irena Kogan und Heike Wirth. Irena Kogan danke ich auch für die Bereitstellung ihres SPSS-Recodierungsprogramms zur Umsetzung von ISEI.

2. Zur Zielsetzung und Konstruktion von ISEI

Die Verbreitung sozioökonomischer Status- und Prestigeskalen geht auf die in den 1960er-Jahren entwickelten pfadanalytischen Modelle zum Staterwerbprozess zurück. In diesen Modellen beeinflusst die soziale Herkunft (Beruf und Bildung des Vaters) den Erwerb von Bildungsqualifikationen des Befragten, die wiederum Einfluss auf den ersten Beruf und den späteren Berufserfolg besitzen. Für die linearen Regressionsanalysen waren quantitative bzw. kontinuierliche Variablen erforderlich. Die Berufsvariablen wurden durch sozioökonomische Statusskalen operationalisiert. Das hinter diesen Skalen stehende Konzept entspricht der Vorstellung, dass Berufe in eine eindimensionale Ordnung gebracht werden können, welche die Stellung der Berufsinhaber in der sozialen Hierarchie ausdrückt. Die Konzentration auf Berufe für die Konstruktion von Statusskalen beruht auf der Einschätzung der Arbeitsteilung als zentralem Element sozialer Ungleichheit.

In dieser Tradition steht auch die Konstruktion von ISEI als Weiterentwicklung früherer sozioökonomischer Indizes für international vergleichende Forschungen. Das verwendete Staterwerbmodell bezieht sich auf drei Kernvariablen und die Beziehungen zwischen ihnen: an zentraler Stelle Einkommen als Maß wirtschaftlichen Wohlstands und Indikator unterschiedlicher Lebensbedingungen sowie Bildung und Beruf als individuelle Ressourcen, die zum Erwerb von Arbeitseinkommen eingesetzt werden.

Die zu skalierenden Berufseffekte werden als zwischen Bildung und Einkommen stehende intervenierende Variable betrachtet. Da sowohl die Verteilung von Bildungsqualifikationen, operationalisiert durch die für den Erwerb eines Bildungsabschlusses typischerweise notwendigen Ausbildungsjahre, als auch die Berufsverteilung mit dem Lebensalter variieren, wurden bei dem Skalierungsmodell die Alterseffekte kontrolliert.

Ganzeboom und Treiman (2003: 171) führen zur Konstruktion von ISEI weiter aus: „We conceive of ISEI as measuring the attributes of occupations that convert a person's education into income. Accordingly, ISEI scores were generated by an optimal scaling procedure in which scores were assigned to occupation unit groups in such a way as to maximise the indirect effect of education on income through occupation and to minimise the direct effect of education on income, net of occupation (with both effects net of age).“

Die Datenbasis bildeten 31 Umfragen zur Schichtung und sozialen Mobilität von 1966 bis 1982 aus 16 verschiedenen Ländern. Für Deutschland wurde die ZUMA-Standarddemographie 1976-1980 (ZA-Studiennummer 1233) benutzt. Für die Analysen wurden vollzeitbeschäftigte männliche Erwerbstätige im Alter von 21 bis 64 Jahren ausgewählt. Diese Begrenzung auf Männer war notwendig, da in der Mehrzahl der größeren Datenbasen keine Angaben über weibliche Erwerbstätige vorlagen. Nach diesen Selektionen

umfasste die Datenbasis rund 74.000 Personen. Ursprünglich wurde die internationale Berufsklassifikation ISCO-68 verwendet (Ganzeboom et al. 1992). Nach einer Umcodierung der Berufsangaben auf ISCO-88 wurden Umsteigeschlüssel bereitgestellt, die eine Recodierung der nach ISCO-88 recodierten Berufsangaben zu ISEI ermöglichen (Ganzeboom/Treiman 1996, 2003).

Neben diesem wichtigen Vorteil der einfachen Recodierung sind an dieser Stelle – bevor im nächsten Abschnitt die Umsetzung mit dem Mikrozensus dargestellt wird – aber auch Einschränkungen festzuhalten. So kann ISEI in der Regel nur für die erwerbstätige Bevölkerung ermittelt werden. Für Nichterwerbstätige ist dies nur möglich, wenn Angaben zu einem früher ausgeübten Beruf vorliegen. Infolge der auf das persönliche Einkommen als abhängige Variable begrenzten Adjustierung bleiben andere Faktoren sozioökonomischer Lagen im Index außer Acht: selbständige vs. abhängige Beschäftigung, befristete Tätigkeiten und geringfügige Beschäftigungsverhältnisse oder Aspekte vertikaler sozialer Ungleichheit, wie beispielsweise die Unterscheidung von Inhabern gleicher Berufe nach Qualifikation, Funktion und Autonomie innerhalb des Betriebes.

Oft ist für die Bestimmung der sozioökonomischen Lage von Personen nicht nur die persönliche Einkommenssituation, sondern auch die gemeinsame Wirtschaftsführung im Haushalt relevant. Haushaltszusammenhänge bleiben aber bei ISEI ebenfalls unberücksichtigt.

Schließlich bezieht sich ein grundsätzlicher methodischer Einwand darauf, dass die einzelnen Komponenten des sozioökonomischen Index unterschiedliche inhaltliche Bedeutungen haben, und deshalb die Zusammenhänge zwischen Einkommen, Bildung und Beruf besser getrennt analysiert werden sollten. Hauser und Warren stellen die Nützlichkeit von Prestige- und Statusskalen grundsätzlich in Frage: „While composite measures of occupational status may have heuristic uses, the global concept of occupational status is scientifically obsolete“ (Hauser/Warren 1997: 251; vgl. auch Rohwer/Pötter 2002: 83ff.). Ganzeboom und Treiman (2003: 161) weisen demgegenüber darauf hin, dass gegenwärtig keine international vergleichbaren Informationen zur beruflichen Bildung und zum Erwerbseinkommen vorhanden sind, sodass schon aus pragmatischen Gründen Skalen wie ISEI bei ländervergleichenden Analysen eingesetzt werden sollten.

3. Die Ausgangsbasis: ISCO-88 (COM) im Mikrozensus ab 1996

Der Internationale Sozioökonomische Index ISEI kann mit den Scientific Use Files des Mikrozensus ab dem Erhebungszeitpunkt 1996 für zwei nach ISCO-88 (COM) recodierte Variablen umgesetzt werden: Der Beruf der ersten Erwerbstätigkeit (EF683) und der Beruf der früheren Erwerbstätigkeit (EF688). Die früheren Scientific Use Files enthalten

keine nach ISCO-88 vercodeten Berufsangaben, sondern nur die nach der nationalen Klassifikation der Berufe (KldB) in der Fassung von 1975 bzw. 1992 verschlüsselten Berufsangaben (vgl. hierzu den Überblick über die Berufsvercodung in Geis/Hoffmeyer-Zlotnik 2000).

Grundsätzlich werden die im Mikrozensus erhobenen Berufsangaben zunächst für die nationale KldB vercodet. Für die Erstellung von ISCO-Codes verwenden die statistischen Ämter einen Umsteigeschlüssel. Bei den Codes für die Berufsuntergruppen (3-Steller) handelt es sich um die für die Verwendung in der Europäischen Gemeinschaft modifizierte Version ISCO-88 (COM). Diese spezielle Version wurde nötig, weil die verschiedenen nationalen Berufsklassifikationen nicht ohne Einschränkungen mit ISCO-88 vergleichbar gemacht werden konnten. Für den Mikrozensus wird die Vergleichbarkeit zwischen ISCO-88 (COM) und der deutschen KldB 1992 – und damit die Güte der Umschlüsselung – als gut eingeschätzt (Elias 1997: 26; vgl. hierzu auch Macht 1992).

Im Unterschied zu ISCO-88 wurde für ISCO-88 (COM) die Berufsuntergruppe 247 „Wissenschaftliche Verwaltungsfachkräfte des öffentlichen Dienstes“ neu geschaffen. Ihr wurde der für die übergeordnete Berufsgruppe 24 „Sonstige Wissenschaftler und verwandte Berufe“ vorgesehene ISEI-Wert zugewiesen. Weitere Unterschiede zwischen ISCO-88 und ISCO-88 (COM) (vgl. hierzu Elias/Birch 1994) sind in Bezug auf die Konstruktion von ISEI vernachlässigbar, so dass die in Ganzeboom/Treiman (2003: 176-191) dargestellte Umschlüsselung von ISCO-88 (3-Steller) zu ISEI übernommen werden kann.

Mit der Variablen EF683 wird die erste Tätigkeit erfasst. Für eine eventuelle zweite Erwerbstätigkeit liegen in den Mikrozensus-Scientific Use Files keine ISCO-88 (COM) Codes vor. Bei Personen, die in der Berichtswoche nicht erwerbstätig waren, werden die Berufsangaben zur früheren Erwerbstätigkeit (EF688) nur in der Unterstichprobe des Mikrozensus erfragt, die mit einem variablen Auswahlatz von durchschnittlich 0,45 Prozent durchgeführt wird. Unter Berücksichtigung fehlender Angaben aufgrund von Nichtbeantwortung liegen in EF688 für rund 25.000 Befragte Berufsangaben vor. Für Soldaten, Wehrpflichtige und Zivildienstleistende können keine ISEI-Werte ermittelt werden.

In den Variablen EF683 und EF688 wurden wenige schwach besetzte Kategorien aus Anonymisierungsgründen mit benachbarten Kategorien beziehungsweise in der Sammelkategorie „Sonstige Arbeitskräfte“ zusammengefasst. Testauswertungen haben gezeigt, dass die ISEI-Verteilungen durch diese Merkmalsvergrößerungen praktisch nicht verändert werden. Die aus der Recodierung der dreistelligen Berufsuntergruppen von ISCO-88 (COM) resultierende Variable ISEI weist ein Minimum von 16 und ein Maximum von 85 auf. Weitere Kennwerte von ISEI werden im nächsten Abschnitt berichtet.

4. Zusammenhänge zwischen ISEI und anderen Variablen im Mikrozensus 2000

Ausgehend von den eingangs genannten Fragen zur Anwendbarkeit des Internationalen Sozioökonomischen Index mit Mikrozensusdaten werden im Folgenden Eigenschaften von ISEI zum gegenwärtigen Beruf beschrieben und insbesondere Zusammenhänge mit dem Einkommen untersucht. Die ISEI-Angaben zum früheren Beruf von Nichterwerbstätigen bleiben hierbei ausser Acht. Ergänzend wird geprüft, in welcher Weise mit ISEI auch Aspekte vertikaler sozialer Ungleichheit beschreibbar sind. Hierfür werden Zusammenhänge zwischen ISEI und der Variablen „Stellung im Betrieb“ dargestellt.

Als Datengrundlage wird das Scientific Use File des Mikrozensus 2000 herangezogen.² Die folgenden Analysen beziehen sich auf die erwerbstätige Bevölkerung am Hauptwohnsitz im Alter von 15-64 Jahren. Da statistische Zusammenhänge zu beschreiben sind, wird auf eine Anpassung der Mikrozensus-Fallzahlen an die Bevölkerungsfortschreibung verzichtet.

4.1 ISEI-Kennwerte für ausgewählte Gruppen

In Tabelle 1 werden Kennziffern der Verteilung von ISEI nach Bundesgebiet, Geschlecht, Staatsangehörigkeit und Arbeitszeit berichtet. Die Verteilungen von ISEI im neuen und früheren Bundesgebiet sowie zwischen Männern und Frauen unterscheiden sich nur geringfügig, wobei die höchsten Werte für vollzeitbeschäftigte Frauen festzustellen sind. Da die ISEI-Werte lediglich eine Recodierung der Angaben von ISCO-88 (COM) darstellen und damit für die betrachteten Gruppen die unterschiedlichen Berufsstrukturen wiedergeben, kann nicht erwartet werden, dass der Index die im Osten Deutschlands im Vergleich zum früheren Bundesgebiet wesentlich niedrigeren Verdienstniveaus widerspiegelt; dies wäre aber bei der Verwendung von ISEI zu beachten.

Größere Unterschiede sind zwischen Deutschen und Migranten festzustellen. Die unterdurchschnittlichen ISEI-Werte für Migranten sind auf deren Konzentration in den Berufshauptgruppen „Handwerksberufe“ sowie „Anlagen- und Maschinenbediener“ zurückzuführen.

Des Weiteren liegen die Kennwerte von ISEI für Teilzeitbeschäftigte niedriger als jene für Vollzeitbeschäftigte, da Teilzeitbeschäftigte häufiger in Büro-, Dienstleistungs- und Hilfs-

2 Für ausführliche Datenbeschreibungen siehe

http://www.gesis.org/Dauerbeobachtung/Mikrodaten/Daten/Abteilungsdaten/Mikrozensus/mz_daten_index.htm#2000.

arbeitsberufen arbeiten. Ähnlich hängen die in den neuen Bundesländern im Vergleich zu Männern höheren ISEI Werte von Frauen mit deren höheren Anteilen in Verwaltungs- und Dienstleistungsberufen zusammen.

Tabelle 1: Kennziffern zu ISEI für ausgewählte Gruppen von Erwerbstätigen

	Mittelwert	1. Dezil	Median	9. Dezil	n
<i>Bundesgebiet</i>					
Männer	44	26	36	69	117.914
Frauen	43	23	43	65	95.635
Deutsche	44	25	39	69	200.061
Migranten	36	20	34	55	13.488
Insgesamt	43	25	39	68	213.549
<i>Früheres Bundesgebiet</i>					
Männer, Teilzeit	43	23	39	69	5.178
Männer, Vollzeit	44	27	38	69	89.684
Zusammen	44	26	38	69	94.862
Frauen, Teilzeit	41	16	39	56	33.099
Frauen, Vollzeit	45	25	43	65	42.100
Zusammen	43	23	43	65	75.199
Insgesamt	44	25	39	68	170.061
<i>Neue Bundesländer</i>					
Männer, Teilzeit	42	23	36	69	1.146
Männer, Vollzeit	42	26	34	69	21.906
Zusammen	42	26	34	69	23.052
Frauen, Teilzeit	41	16	43	65	5.410
Frauen, Vollzeit	46	25	47	68	15.026
Zusammen	45	25	43	67	20.436
Insgesamt	43	25	39	68	43.488

Quelle: Mikrozensus 2000 (faktisch anonymisierte 70%-Substichprobe). Datenkonstruktion/Fallauswahl: Erwerbstätige Bevölkerung im Alter von 15-64 Jahren am Hauptwohnsitz; ungewichtete Fallzahlen; Teil-/Vollzeit: unter 35 bzw. 35 und mehr normalerweise geleistete Wochenarbeitsstunden. ISEI-Wertebereich: 16-85.

Wie im zweiten Abschnitt erwähnt, wurden die ISEI-Werte nur für männliche Erwerbstätige ermittelt. Ganzeboom und Treiman (1996: 218) sprechen sich dennoch aus konzeptionellen Gründen für eine Übertragung von ISEI für Frauen aus. Dies scheint aber, wie in Tabelle 1 zu erkennen, nicht für alle Anwendungsfälle uneingeschränkt möglich. Beispielsweise würde eine Darstellung der geschlechtsspezifischen beruflichen Ungleichheit mithilfe des ISEI den Eindruck geringer Unterschiede vermitteln. Jedoch liegen die Einkommen vollzeiterwerbstätiger Frauen im früheren Bundesgebiet nach Ergebnissen auf Basis von Mikrozensusdaten 2000 lediglich bei rund 65 Prozent des monatlichen Nettoeinkommens von Männern (siehe Abschnitt 4.3). Nach Ergebnissen der Lohn- und Gehaltsstrukturerhebung 2001 betragen die Verdienstabstände beim Bruttomonatsverdienst im Bundesgebiet für die beiden höchsten Leistungsgruppen von Angestellten – also selbst bei vergleichbaren beruflichen Tätigkeiten und unter Kontrolle von Leistungsmerkmalen – 76 bzw. 83 Prozent (Frank-Bosch 2003: 29). Vor diesem Hintergrund erscheint die Übertragbarkeit und Anwendung von ISEI für Frauen nur unter Vorbehalten und Verwendung weiterer Merkmale gerechtfertigt.

4.2 Stellung im Betrieb

Eine Möglichkeit, die ausgeübten Berufe nach leistungs- und tätigkeitsbezogenen Kriterien zu unterscheiden, bietet die Variable „Stellung im Betrieb“, in der die sozialrechtlichen Gruppen der Erwerbstätigen im Mikrozensus weiter differenziert werden: Selbständige werden nach der Zahl ihrer Mitarbeiter unterschieden, Beamte nach ihrem Dienstverhältnis, Angestellte nach Qualifikation, Aufsichts- und Dispositionsumfang und Arbeiter nach beruflicher Qualifikation und Aufsichtsfunktion.

Tabelle 2 zeigt, dass Beamte im höheren und gehobenen Dienst, gefolgt von den beiden oberen Gruppen der Angestellten, durchschnittlich die höchsten, an- und ungelernete Arbeiter die niedrigsten ISEI-Werte aufweisen. Selbständige mit fünf und mehr Beschäftigten nehmen erst den vierten Rang ein. Beamte im einfachen und mittleren Dienst rangieren höher als Angestellte, die schwierige Aufgaben nach allgemeinen Vorgaben selbständig ausführen (z.B. Buchhalter). Auch stehen Beamte des einfachen und mittleren Dienstes im Rang nicht nur besser als Meister im Angestelltenverhältnis, sondern liegen auch weit vor Meistern im Arbeiterverhältnis.

Tabelle 2: ISEI-Werte nach der betrieblichen Stellung der Erwerbstätigen (Männer und Frauen)

Betriebliche Stellung	Mittelwert	1. Dezil	Median	9. Dezil	n
Auszubildende, Praktikanten, Volontäre	39	29	36	55	9.291
Selbständige					
mit 1 bis 4 Beschäftigten	49	23	51	73	17.170
mit 5 und mehr Beschäftigten	54	30	51	85	3.718
Mithelfende Familienangehörige	31	16	30	51	1.951
Beamte					
im einfachen Dienst	49	39	47	69	593
im mittleren Dienst	53	40	51	69	3.748
im gehobenen Dienst	63	47	68	69	4.759
im höheren Dienst	70	65	69	85	2.658
Arbeiter					
An- und ungelernete Arbeiter	28	16	29	39	30.224
Facharbeiter, Gesellen	33	26	34	40	31.420
Vorarbeiter, Kolonnenführer	34	23	34	48	3.732
Meister, Poliere im Arbeiterverhältnis	36	29	34	49	1.043
Angestellte					
Meister, Poliere im Angestelltenverhältnis	43	30	38	67	3.073
Ausführende Angestellte	43	27	43	55	10.089
mit einfachen Fachtätigkeiten	44	32	43	55	20.080
mit schwierigen Aufgaben	47	34	49	65	36.452
selbst. Leistung in verantwortl. Tätigk.	58	38	56	73	23.709
mit umfassenden Führungsaufgaben	58	39	55	73	5.238
Zusammen	44	25	39	68	208.948

Quelle: Mikrozensus 2000 (faktisch anonymisierte 70%-Substichprobe); Datenkonstruktion/Fallauswahl: siehe Tab. 1; gültige Angaben zu ISEI und zur betrieblichen Stellung. ISEI-Wertebereich: 16-85.

Mit Ausnahme der beiden oberen Angestelltengruppen variiert ISEI innerhalb der einzelnen sozialrechtlichen Gruppen mit der Stellung in der betrieblichen Hierarchie, jedoch scheint diese Binnendifferenzierung in Bezug auf die typischen Einkommenschancen relativ gering zu sein.

Zur Überprüfung dieser Frage werden Verhältniswerte der Einkommen und von ISEI für einzelne Gruppen berechnet. Die Verhältnisse der Einkommen betragen bei höheren vs. einfachen Beamten (5.680 / 3.130 DM) und bei den beiden Arbeitergruppen (3.276 / 1.870 DM) rund das 1,8-fache, bei den Angestellten in Führungspositionen vs. ausführenden Angestellten (6.730 / 1.820 DM) sogar das 3,7-fache. Dagegen liegen die Verhältnisse der ISEI-Werte für höhere vs. einfache Beamte ($70 / 49 = 1,4$), Meister im Arbeiterverhältnis vs. an- und ungelernte Arbeiter (1,3) sowie Angestellte in Führungspositionen vs. ausführende Angestellte (1,3) deutlich unter den für diese Gruppen berechneten Verhältniswerten der mittleren Monatsnettoeinkommen.

Zusammenfassend lässt sich festhalten, dass bei ISEI den Berufen Werte ohne Berücksichtigung der ungleichen sozialstrukturellen Lage zugewiesen werden. Die nach der betrieblichen Stellung differenzierten ISEI-Werte sowie die beispielhaften Angaben zur Einkommensschichtung deuten darauf hin, dass mit ISEI nur eingeschränkt vertikale Aspekte sozialer Ungleichheit abgebildet werden können.

4.3 Geschlechtsspezifische Einkommen

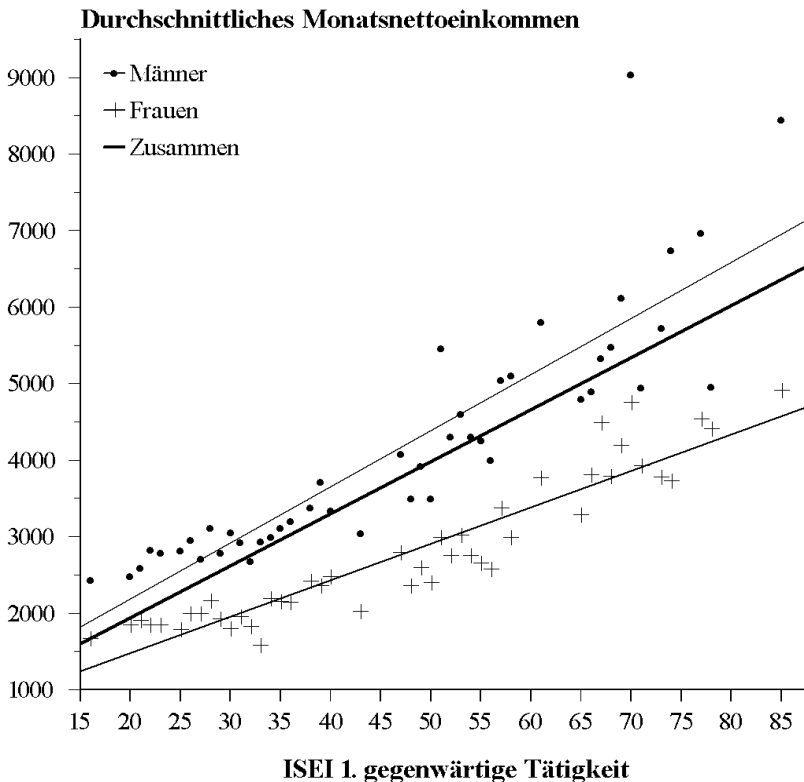
In der amtlichen Statistik dienen die Einkommensangaben des Mikrozensus vorwiegend zur ergänzenden Beschreibung der sozioökonomischen Lage von Personen und Haushalten. Bei der Frage nach dem individuellen Monatsnettoeinkommen werden neben Einkünften aus der Erwerbstätigkeit alle weiteren Einkommensquellen erfasst; unter anderem öffentliche Unterstützungen oder Einkommen aus Vermietung und Verpachtung. Bei selbständigen Landwirten entfällt die Angabe des Einkommens. Die Angaben liegen gruppiert in 24 Kategorien mit einer nach oben offenen Randklasse (35.000 DM oder mehr) vor.

Um die Einkommensangaben näherungsweise auf Erwerbseinkommen beziehen zu können, werden im Folgenden nur Erwerbstätige ausgewählt, die ihren überwiegenden Lebensunterhalt aus Erwerbstätigkeit erzielen. Um weiterhin die Heterogenität nicht kontrollierter Faktoren einzuschränken, werden nur Vollzeit-erwerbstätige im früheren Bundesgebiet mit 35 und mehr Stunden normalerweise geleisteter wöchentlicher Arbeitszeit betrachtet. Für die Einkommensberechnung werden Gruppenmittelwerte verwendet. Die nach oben offene Randklasse erhält den 1,5-fachen Wert der Kategorienuntergrenze zugewiesen (vgl. Cramer 1986: 241).

Aufgrund der Konstruktion von ISEI ist ein enger Zusammenhang zwischen dem Einkommen und dem sozioökonomischen Index zu erwarten. Abbildung 1 zeigt, dass Frauen mit durchschnittlich 2.590 DM rund 65 Prozent des Einkommens von Männern (3.970 DM) erreichen; der Gesamtdurchschnitt liegt bei 3.530 DM. Wie anhand der Reg-

ressionsgeraden zu erkennen ist, verläuft der Zusammenhang zwischen ISEI und dem Einkommen bei Frauen flacher als bei Männern.³ Wie bereits im Abschnitt 4.1 erwähnt, befürworten Ganzeboom und Treiman (1996: 218) die Übertragung von ISEI-Werten für Frauen. Diese Ergebnisse zeigen dagegen, welche Schwierigkeiten damit verbunden sind.

Abbildung 1: Durchschnittliches Monatsnettoeinkommen nach Geschlecht und ISEI



3 Allerdings ist hierbei zu beachten, dass das Nettoeinkommen vom Steuer- und Transfersystem abhängt und verheiratete Frauen häufiger in einer ungünstigeren Steuerklasse eingestuft sind.

Trotz der vorgenommenen Fallauswahl können weitere Einkommensarten und einkommensrelevante Sachverhalte eine Rolle spielen. Nicht zuletzt ist zu berücksichtigen, dass die Einkommen erheblich stärker als in Abbildung 1 streuen, da hier lediglich gruppenspezifische Mittelwerte dargestellt sind. Zur Kontrolle dieser und weiterer Effekte wird im nächsten Abschnitt eine lineare Regression des Einkommens durchgeführt.

4.4 Lineare Regression des Monatsnettoeinkommens

Bei Regressionsanalysen des Monatsnettoeinkommens mit Mikrozensusdaten sind zur Berücksichtigung der Effekte des Steuer- und Transfersystems besondere Annahmen, Selektionen und Kontrollvariablen nötig (vgl. zusammenfassend Schimpl-Neimanns 2002).

Die folgenden Analysen konzentrieren sich auf männliche abhängig beschäftigte Erwerbstätige im Alter von 15 bis 64 Jahren im früheren Bundesgebiet, die in der Berichtswoche regelmäßig wenigstens 35 Stunden (Vollzeitbeschäftigte) gearbeitet haben und deren überwiegende Quelle des Lebensunterhalts die Erwerbstätigkeit darstellt. Ausgeschlossen sind Auszubildende und Personen mit einer zweiten Erwerbstätigkeit sowie Fälle mit fehlenden Angaben in den verwendeten Variablen. Darüber hinaus beschränkt sich die Fallauswahl auf Personen, die an der in den Mikrozensus integrierten EU-Arbeitskräftestichprobe teilgenommen haben, da einkommensrelevante Angaben über besondere Arbeitsformen (regelmäßige oder ständige Samstags-, Sonntags-, Feiertags- oder Schichtarbeit) nur in der Substichprobe vorliegen. Die Einkommensangaben werden, wie in den obigen Analysen, auf die jeweilige Klassenmitte gesetzt und zusätzlich logarithmiert.⁴

Der Zusammenhang zwischen ISEI und dem Einkommen beträgt für diese Teilstichprobe und ohne Kontrollvariablen $R^2 = 0,241$ ($r = 0,49$). Das erste Modell (s. Tab. 3) enthält in Anlehnung an humankapitaltheoretisch begründete Einkommensfunktionen verschiedene einkommensrelevante Variablen. Neben dem Alter als Proxy-Variablen für die Berufserfahrung wird die Wochenarbeitszeit, das Vorliegen besonderer Arbeitsformen und die Angabe, ob neben dem Erwerbseinkommen zusätzliche Einkommensquellen vorhanden sind, berücksichtigt. In Bezug auf Effekte des Steuer- und Transfersystems werden für Familienvorstände die Zahl nichterwerbstätiger lediger Kinder unter 27 Jahren in der Familie und die Angabe, ob die Ehefrau Erwerbseinkommen besitzt, als Kontrollvariablen verwendet. Die allgemeinbildende und berufliche Bildungsqualifikation wird mithilfe der CASMIN-Bildungsklassifikation (Brauns et al. 2000, Granato 2000) umgesetzt. Dieses erste Modell, in dem ISEI nicht enthalten ist, erklärt statistisch 41 Prozent der Einkommensvarianz ($R^2 = 0,409$).

4 Die Schätzwerte der linearen Regression unterscheiden sich nur unwesentlich von den Ergebnissen von Regressionsmodellen, mit denen die Gruppierung und Zensierung der Einkommensangaben statistisch korrekt behandelt wird.

Tabelle 3: Lineare Regression des logarithmierten Monatsnettoeinkommens erwerbstätiger Männer im früheren Bundesgebiet

Variablen	Modell 1		Modell 2		Modell 3		Modell 4	
	B	s.e.	B	s.e.	B	s.e.	B	s.e.
Konstante	5,855	(0,069)	5,785	(0,067)	6,534	(0,065)	6,663	(0,065)
Alter	0,020	(0,002)	0,019	(0,002)	0,018	(0,001)	0,018	(0,001)
Alter ² / 100	-0,011	(0,002)	-0,011	(0,002)	-0,011	(0,002)	-0,011	(0,002)
Wochenarb.-zeit (ln (Std.))	0,392	(0,016)	0,347	(0,016)	0,194	(0,016)	0,193	(0,016)
Besondere Arbeitsformen	0,002	(0,005)	0,018	(0,005)	0,028	(0,004)	0,024	(0,004)
Zusatz Einkommen	0,128	(0,005)	0,119	(0,005)	0,106	(0,005)	0,108	(0,005)
nicht Familienvorstand	-0,135	(0,008)	-0,130	(0,008)	-0,124	(0,007)	-0,124	(0,007)
Zahl Kinder in Familie	0,069	(0,003)	0,072	(0,003)	0,071	(0,002)	0,071	(0,002)
Ehefrau mit Erwerbseink.	-0,022	(0,004)	-0,021	(0,004)	-0,024	(0,004)	-0,025	(0,004)
CASMIN-Bildungsklassifik.								
1a Kein Abschluss	-0,345	(0,017)	-0,292	(0,017)	-0,223	(0,016)	-0,229	(0,017)
1b Hauptschule	-0,167	(0,008)	-0,123	(0,008)	-0,066	(0,008)	-0,072	(0,008)
2b Mittlere Reife	-0,108	(0,019)	-0,120	(0,018)	-0,093	(0,017)	-0,091	(0,017)
2a Mittlere Reife + Berufsausb	0,147	(0,005)	0,089	(0,005)	0,051	(0,005)	0,057	(0,005)
2c_gen Abitur	0,011	(0,022)	-0,069	(0,022)	-0,078	(0,021)	-0,061	(0,021)
2c_voc Abitur + Berufsausb.	0,227	(0,008)	0,123	(0,008)	0,064	(0,008)	0,080	(0,008)
3a Fachhochschulabschluss	0,402	(0,008)	0,198	(0,009)	0,125	(0,009)	0,178	(0,008)
3b Hochschulabschluss	0,484	(0,008)	0,256	(0,009)	0,168	(0,009)	0,231	(0,009)
ISEI 1. gegenw. Erwerbstätigk.			0,008	(0,000)	0,004	(0,000)		
Betriebliche Stellung								
Beamte im einfachen Dienst					0,006	(0,028)	0,049	(0,028)
... im mittleren Dienst					0,116	(0,012)	0,176	(0,012)
... im gehobenen Dienst					0,215	(0,012)	0,288	(0,012)
... im höheren Dienst					0,310	(0,016)	0,394	(0,016)
Arbeiter: An- und Ungelernte					-0,103	(0,007)	-0,112	(0,007)
... Vorarbeiter, Kolonnenführ.					0,041	(0,011)	0,046	(0,011)
... Meister, Poliere					0,083	(0,019)	0,091	(0,019)
Angestellte: Meister, Poliere					0,126	(0,012)	0,163	(0,011)
... Ausführende					-0,096	(0,013)	-0,068	(0,013)
... einfache Fachtätigkeiten					-0,085	(0,010)	-0,050	(0,010)
... schwierige Aufgaben					0,048	(0,007)	0,097	(0,007)
... in verantwortlicher Tätigk.					0,202	(0,008)	0,273	(0,007)
...umfass. Führungsaufgaben					0,493	(0,011)	0,567	(0,011)
Korrigiertes R ²	0,409		0,446		0,499		0,492	

Quelle: Mikrozensus 2000 (faktisch anonymisierte 70%-Substichprobe).

Auswahl der Teilstichprobe (n=28.781): EU-Arbeitskräftestichprobe; Alter: 15-64 Jahre; Bevölkerung am Ort der Hauptwohnung; in der Berichtswoche regelmäßig Erwerbstätige mit überwiegendem Lebensunterhalt aus Erwerbstätigkeit; normalerweise geleistete Arbeitszeit: 35 und mehr Stunden/Berichtswoche; abhängig Beschäftigte (Beamte, Arbeiter und Angestellte) ohne zweite Erwerbstätigkeit; gültige Angaben zu allen verwendeten Variablen.

Referenzkategorie: keine besondere Arbeitsform; keine Zusatz Einkommen; Familienvorstand; Ehefrau des Familienvorstands ohne Erwerbseinkommen; CASMIN-Bildungsklassifik.: Hauptschule + Berufsausbildung. (1c); Betriebliche Stellung: Facharbeiter.

Wird ISEI im zweiten Modell hinzugefügt, steigt die Anpassungsgüte um vier Prozent auf $R^2 = 0,446$. Somit kann unter Kontrolle der auch im ersten Modell verwendeten Variablen mit ISEI nur ein relativ geringer Varianzanteil des Einkommens statistisch erklärt werden. Während die Koeffizienten, wie beispielsweise zum Alter, in den verschiedenen Modellen weitgehend gleich bleiben, ist die Aufnahme von ISEI aufgrund des engen Zusammenhangs zwischen ISEI und der Bildungsqualifikation vor allem mit Veränderungen der Koeffizienten zur CASMIN-Bildungsklassifikation verbunden.

Wie der Vergleich der Modelle 2 und 3 zeigt, lässt sich mit der zusätzlichen Aufnahme der Variablen „Stellung im Betrieb“ die Anpassungsgüte des Modells um fünf Prozent auf $R^2 = 0,499$ steigern. Der Ausschluss von ISEI im Modell 4 reduziert dagegen das R^2 um lediglich 0,7 Prozent auf $R^2 = 0,492$. Damit besitzt ISEI eine zwar statistisch signifikante, aber doch insgesamt sehr geringe Erklärungskraft und fällt deutlich hinter die Variable „Stellung im Betrieb“ zurück.⁵ Allerdings werden diese Angaben zur Stellung im Betrieb gegenwärtig im Mikrozensus nur im vierjährigen Abstand (1996, 2000, 2004) erhoben. Sie sind darüber hinaus auch nicht direkt für ländervergleichende Analysen verwendbar und können deshalb ISEI oder andere international vergleichbare Klassifikationen nicht ersetzen.

5. Schluss

Mit der Konstruktion von ISEI wurde das Ziel verfolgt, die unterschiedlichen Eigenschaften von Berufen zu erfassen, mit denen Erwerbstätige Einkommen erzielen. Da in Umfragedaten inzwischen in den meisten Ländern nach ISCO-88 verschlüsselte Berufsangaben vorliegen, kann ISEI sehr einfach erstellt werden. Darüber hinaus erlaubt ISEI als eine kontinuierliche Variable in statistischen Analysen sparsamere Modelle als sozioökonomische Klassifikationen, die – wie beispielsweise das EGP-Klassenschema (vgl. Brauns et al. 2000) – als qualitative Variable zu behandeln sind. Diese Vorteile haben wesentlich zur breiten Verwendung von ISEI in der ländervergleichenden Forschung beigetragen. Mit dem Vorteil der einfachen Konstruktion von ISEI durch die Recodierung von ISCO-88 sind aber auch Probleme verbunden.

In Bezug auf die erheblichen Einkommensunterschiede zwischen Männern und Frauen, die bei der Konstruktion von ISEI auf Basis internationaler Daten ausschließlich von männlichen Erwerbstätigen unberücksichtigt blieben, ist die Übertragung und Anwendung

5 Der Zusammenhang zwischen dem logarithmierten Monatsnettoeinkommen und der Variablen „Betriebliche Stellung“ in dieser Substichprobe beträgt ohne weitere Kontrollvariablen $R^2 = 0,341$ und ist stärker als der Zusammenhang zwischen dem Einkommen und ISEI ($R^2 = 0,241$).

von ISEI für Frauen nur eingeschränkt in der Lage, die unterschiedlichen geschlechtsspezifischen Ungleichheiten wiederzugeben. Dieses Problem teilt ISEI mit anderen sozio-ökonomischen Skalen und Klassifikationen (vgl. Hauser/Warren 1997: 198f.)

Die Ergebnisse der Regressionsanalyse deuten darauf hin, dass die berufsspezifischen Einkommenseffekte im Vergleich zu den weiteren verwendeten Variablen zur Arbeitssituation, der Bildungsqualifikation und der betrieblichen Stellung eher schwach ausgeprägt sind. Insbesondere mit den in der Variablen „Betriebliche Stellung“ erfassten leistungs- und tätigkeitsbezogenen Sachverhalten können die für den deutschen Arbeitsmarkt spezifischen sozialstrukturellen Zusammenhänge wesentlich besser erklärt werden als mit ISEI.

Diese Punkte sprechen nicht gegen den Einsatz von ISEI in ländervergleichenden Analysen, zumal die Korrelation zwischen ISEI und dem Monatsnettoeinkommen ohne weitere Kontrollvariablen $r = 0,49$ beträgt. Die hier für den Mikrozensus 2000 berichteten Ergebnisse können aber dazu beitragen, die Möglichkeiten und Einschränkungen bei der Verwendung von ISEI mit Mikrozensusdaten besser zu beurteilen.

Literatur

Baumert, J./Klieme, E./Neubrand, M./Prenzel, M./Schiefele, U./Schneider, W./Stanat, P./Tillmann, K.-J./Weiß, M., (Hrsg.): PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske + Budrich.

Brauns, H./Steinmann, S./Haun, D., 2000: Die Konstruktion des Klassenschemas nach Erikson, Goldthorpe und Portocarero (EGP) am Beispiel nationaler Datenquellen aus Deutschland, Großbritannien und Frankreich. ZUMA-Nachrichten 46: 7-63.

Cramer, J.S., 1986: Estimation of Probability Models From Income Class Data. *Statistica Neerlandica* 40(4): 237-250.

Elias, P., 1997: Occupational Classification: concepts, methods, reliability, validity and cross-national comparability. Luxemburg: LES Working Paper 5. <URL: <http://www.lisproject.org/publications/leswps/leswp5.pdf>>

Elias, P./Birch, M., 1994: Errichtung einer EG-weiten Statistik der Berufe. ISCO 88 (COM). Fassung der Internationalen Standardklassifikation der Berufe 1988 zur Verwendung innerhalb der Europäischen Gemeinschaft. Warwick: Institute for Employment Research. <URL: <http://www.warwick.ac.uk/ier/isco/isco88.html>>

Frank-Bosch, B., 2003: Löhne und Gehälter in Deutschland. Gehalts- und Lohnstruktur-erhebung 2001. Hrsg.: Statistisches Bundesamt. Wiesbaden. <URL: http://www.destatis.de/presse/deutsch/pk/2003/loehne_gehaelter_2001_neu.pdf>

Ganzeboom, H.B.G. (WWW-Site): Tools for deriving status measures from ISKO-88. <http://home.scw.vu.nl/~ganzeboom/ismf/scaleapp.htm>.

Ganzeboom, H.B.G./de Graaf, P.M./Treiman, D.J., 1992: A Standard International Socio-Economic Index of Occupational Status. *Social Science Research* 21(1): 1-56.

Ganzeboom, H.B.G./Treiman, D.J., 1996: Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations. *Social Science Research* 25: 201-239.

Ganzeboom, H.B.G./Treiman, D.J., 2003: Three Internationally Standardised Measures for Comparative Research on Occupational Status. S. 159-193 in: J.H.P. Hoffmeyer-Zlotnik/C. Wolf (Hrsg.): *Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables*. New York: Kluwer Academic / Plenum Publishers.

Geis, A.J./Hoffmeyer-Zlotnik, J.H. P., 2000: Stand der Berufsvercodung. *ZUMA-Nachrichten* 47: 103-128.

Granato, N., 2000: Mikrodaten-Tools. CASMIN-Bildungsklassifikation. Eine Umsetzung mit dem Mikrozensus 1996. *ZUMA-Technischer Bericht* 2000/12.

Handl, J./Steinmetz, S., 2003: Geschlechtsspezifische berufliche Segregation in den Ländern Europas: Methodische Probleme und inhaltliche Ergebnisse. Beitrag zur 3. Nutzerkonferenz: „Forschung mit dem Mikrozensus: Analysen zur Sozialstruktur und zum Arbeitsmarkt“, Mannheim, 9.-10. Okt. 2003. <URL: http://www.gesis.org/Dauerbeobachtung/Mikrodaten/documents/Veranstaltungen/Nutzerkonferenz2003/paper/text_handl_steinmetz.pdf>

Hauser, R.M./Warren, J.R., 1997: Socioeconomic Indexes for Occupations: A Review, Update, and Critique. S. 177-298 in: A.E. Raftery (Hrsg.): *Sociological Methodology* 1997. Washington D.C. : The American Sociological Association.

Kogan, I./Müller, W. (Hrsg.), 2003: *School-to-Work Transitions in Europe: Analyses of the EU LFS 2000 Ad Hoc Module*. Mannheim: Mannheimer Zentrum für Europäische Sozialforschung.

Macht, A., 1992: Klassifizierung der Berufe 1992. *Wirtschaft und Statistik* (12): 855-863.

Müller, W./Gangl, M. (Hrsg.), 2003: Transitions from Education to Work in Europe. The Integration of Youth into EU Labour Markets. Oxford: University Press.

Rohwer, G./Pötter, U., 2002: Methoden sozialwissenschaftlicher Datenkonstruktion. Weinheim: Juventa.

Schimpl-Neimanns, B., 2002: Anwendungen und Erfahrungen mit dem Scientific Use File des Mikrozensus. ZUMA-Arbeitsbericht 2002/01. Mannheim: ZUMA.

Korrespondenzadresse

Bernhard Schimpl-Neimanns

ZUMA

Postfach 12 21 55

D – 68072 Mannheim

email: schimpl-neimanns@zuma-mannheim.de

BUCHBESPRECHUNGEN

Es muss schon ein besonderes deutsches Statistikbuch sein, das in der elften Auflage nach alten Rechtschreibregeln erscheint und bereits in englischer, russischer und spanischer Übersetzung vorliegt. Im Rahmen des iChemLab-Projekts (http://ebooks.tuwien.ac.at/common/elearningtag_20040120.pdf) ist die 1999er Auflage sogar als elektronische Kopie vorhanden.

SACHS, LOTHAR

Angewandte Statistik

Berlin: Springer Verlag, 2004

ISBN 3-540-40555-0 11. Auflage, 889 Seiten, 49,95 €

Als Arbeitshilfen sind im Buch 91 Abbildungen und grafische Darstellungen, 888 Formeln und mathematisch-statistische Sätze, 317 Tabellen und 99 Übersichten zu finden. 542 durchgerechnete Beispiele erleichtern den Umgang mit der Theorie. Das Literaturverzeichnis umfasst die Seiten 691-735 (hier ist auf S. 690 einer der wenigen Schreibfehler) und ist mit etwa 1000 Literaturangaben enorm groß und aktuell. Eine Übersicht einiger Symbole und eine kleine Auswahl englischer Fachausdrücke und Abkürzungen werden gegeben. Das Sachverzeichnis von ca. 100 zweiseitigen Seiten liefert eine Fülle von Nachschlagmöglichkeiten und soll dazu dienen Leserinnen und Leser in das Buch „hineinzuziehen“. Man findet dort nicht nur die statistischen Bezeichnungen, sondern etwa unter Äpfel mit Flecken auch einen Verweis auf ein Beispiel zum Multiplikationssatz der Wahrscheinlichkeitsrechnung. Beim Konfidenzintervall wird man auf das deutsche Wort Vertrauensintervall geleitet. Bei Begriffen wie Nonresponse werden auch Seiten im Literaturverzeichnis aufgeführt. Kein Wunder also, dass dieses Buch auch bei vielen Literaturempfehlungen statistischer Vorlesungen vorkommt. Es eignet sich eben „zum Lernen, Anwenden und Nachschlagen für anwendungsorientierte Leser mit unterschiedlichen Vorkenntnissen und breit gestreuten Interessen“, wie der Autor im Vorwort zur achten Auflage schreibt. Am Ende des Buches ist ein Lesezeichen mit Hinweisen eingefügt, die helfen sollen, „den Inhalt des Buches zu erschließen“.

Der Autor gibt mit seinem Buch einen Grundriss der Statistik als Entscheidungshilfe und ist auch für mathematisch nicht vorgebildete Leser zugänglich. Das Buch ist in sieben Kapitel unterteilt. Im ersten Kapitel werden die Grundlagen und Aufgaben der Statistik behandelt, die Elemente der Wahrscheinlichkeitsrechnung erläutert und der Weg zur Normalverteilung besprochen. Die Bausteine eines statistischen Tests werden erläutert und die in der Anwendung wichtigen χ^2 -, t - und F -Verteilung abgehandelt. Bei den diskreten Verteilungen findet man die binomiale, hypergeometrische und Poisson Verteilung sowie als Verallgemeinerung die Multinomialverteilung. Das zweite Kapitel ist verhältnismäßig kurz und referiert über die Epidemiologie. Für den Sozialwissenschaftler wären hier natürlich andere als medizinische Beispiele interessanter. Im dritten Kapitel werden Ein- und Zweistichprobentests vorgestellt, aber auch Ausreißerprobleme behandelt. Dabei wurden auch verteilungsunabhängige Verfahren nicht vergessen. Das vierte Kapitel widmet sich weiteren Prüfverfahren wie etwa dem Vierfeldertest, dem χ^2 -Anpassungstest und dem Kolmogoroff-Smirnoff-Test für die Güte der Anpassung. Regression und Korrelation sind im fünften Kapitel zu finden. Der Auswertung von Mehrwegtafeln im sechsten Kapitel folgt die Varianzanalyse im letzten Kapitel.

Was könnte an dem Buch noch verbessert werden? Natürlich könnten weitere statistische Verfahren dazukommen, etwa die multivariaten Verfahren. Das würde aber den Umfang des Buches sprengen. Bleibt also vielleicht die Lesbarkeit anzusprechen, obwohl hier schon viel Vorbildliches etwa durch Übersichtlichkeit der Gliederung, Markieren wichtiger Punkte durch Ausrufezeichen am Rand usw. geschehen ist. Vergleicht man aber Tabellen miteinander, so stellt man fest, dass die Einheitlichkeit im Schrifttyp fehlt, als wenn Tabellen aus früheren Zeiten mit Courier einfach kopiert und mit anderen Schriftarten vermischt wurden. Man vergleiche etwa Tabelle 4 oder Abbildung 4, wo zum Teil noch handschriftlich Mengenoperationen vorkommen.

Für den Sozialwissenschaftler steht das vorliegende Buch in Konkurrenz zum Springer-Lehrbuch von Bortz, Statistik für Sozialwissenschaftler, das eine ähnliche Seitenzahl hat, in vielen Teilen allerdings weit weniger ausführlich ist, dafür aber multivariate Methoden beinhaltet. Als Fazit werde ich zukünftig das Buch von Sachs bei Anfragen zu statistischen Fragen vermehrt empfehlen, da es aus meiner Sicht eine klare Sprache hat, gut verständlich ist und weite Teile der Statistik abdeckt.

SPSS ist eines der am meist genutzten und umfangreichsten Statistikprogramme für die Sozialwissenschaften. Neben den ausführlichen Programmbeschreibungen von SPSS selbst gibt es einhergehend mit der raschen Entwicklung des Programmsystems und seiner relativ einfachen Nutzung unter Windows eine Vielzahl an Einführungen und Arbeitsbüchern, die vor allem die methodische Ausbildung von Studenten begleiten und vertiefen soll. Datenanalyse mit SPSS für Windows ist bereits in der 3. Auflage neubearbeitet erschienen.

WITTENBERG, REINHARD & CRAMER, HANS

Datenanalyse mit SPSS für Windows.

Handbuch für computerunterstützte Datenanalyse, Band IX.

3. neubearbeitete Auflage

Stuttgart: Lucius & Lucius (UTB), 2003

ISBN 3-8282-0266-7, 324 Seiten, 89 Abbildungen, 18,90 €

Es ist als Lehr- und Arbeitsbuch konzipiert, baut auf einem Grundlagenband des Erstautors auf und ist das alles in allem das gelungene Ergebnis jahrelanger Erfahrung aus der universitären Ausbildung in sozialwissenschaftlichen Methoden und computerunterstützter Datenerfassung und -analyse. Voraussetzung für ein sinnvolles Arbeiten mit dem Handbuch sind Kenntnisse in Theorien und Methoden der computerunterstützten Datenanalyse. Dass das Buch nicht erschöpfend auf alle Möglichkeiten von SPSS (in seiner deutschen Version 11.5) eingehen kann und will, bedarf keiner Erklärung. Zielsetzung von Reinhard Wittenberg und Hans Cramer ist es, häufig benutzte Verfahren der explorativen, deskriptiven und konfirmativen Datenanalyse detailliert zu präsentieren und zu erörtern. Darüber hinaus werden einige Möglichkeiten des Datenmanagements (Datenselektion, -modifikation, und -transformation) behandelt. Eine Besonderheit des vorliegenden Bandes ist, dass die ausgewählten Möglichkeiten von Datenmanagement und Analyseverfahren nicht nur menügesteuert vorgestellt werden, sondern gleichzeitig auch das jeweilige Pendant in SPSS-Befehlssyntax vorgeführt wird. SPSS bietet über seine Syntax nach wie vor mehr als über seine Menüs; Syntax kann als File gespeichert, geändert, korrigiert und weiterverwendet werden; und insbesondere bei umfangreichem Datenmanagement ist die Nutzung der SPSS-Syntax unumgänglich, um eine Dokumentation der Datenbearbeitung zu gewährleisten. Diese Doppelstrategie, die geschickt fast über das ganze Buch verfolgt wird, ist ein deutlicher Pluspunkt im Vergleich zu anderen Einführungstexten. Bei der Erläuterung der einzelnen Statistikprozeduren wird auf die Voraussetzungen, die an die Daten gestellt werden sollten, hingewiesen. Nützlich und hilfreich in einem Lehrbuch sind auch die Interpretationen der Analyseresultate. Ein weiteres Plus ist die Verwendung

eines der in Deutschland meist genutzten sozialwissenschaftlichen Datensätze (oder wegen seines Umfangs eines Teiles davon), der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) aus dem Jahre 2000, in der ein repräsentativer Querschnitt der Bevölkerung Deutschlands zu sozialwissenschaftlichen Themenbereichen befragt wurde. Dadurch bleiben Datenmanagement und -analysen nicht auf einer formalen Ebene stehen, sondern werden praxisnah und unter forschungsrealistischeren Bedingungen vorgeführt, als dies ein auf wenige Fälle und Variablen beschränkter experimenteller Datensatz bieten könnte. Gut im Rahmen eines solchen Lehr- und Arbeitsbuches sind auch die Übungsaufgaben (mit Auflösung im Anhang), die den jeweiligen Kapitelinhalt vertiefen. Ein Minus ergibt sich durch die mangelnde Sorgfalt, die gelegentlich der Befehlssyntax gewidmet wurde (siehe hierzu Fußnoten 4, 6, 7, 8 und 10 dieser Buchbesprechung).

Die Autoren teilen ihr Buch in 5 Kapitel (ein sechstes als Anhang) ein und beginnen in der Einleitung mit den Modulen des Programmsystems und deren Inhalten. Diese Information ist aus Kostengründen durchaus von Relevanz, da die Anschaffungskosten von den benötigten Modulen abhängen.

Im zweiten Kapitel, das ca. ein Drittel des Buches umfasst, werden zunächst die Grundlagen zum Arbeiten mit SPSS für Windows beschrieben: Wie wird das Programm gestartet und beendet; welche Fenster und Editoren stehen zur Verfügung; welche Programmoptionen sollten eingestellt werden¹; wie wird gedruckt und zuvor das Seitenlayout überprüft und geändert? Die Grundlagen von Datenmatrix, Rohdaten und SPSS-Datendateien werden dargestellt. Wie die SPSS-Befehlssprache aufgebaut ist, welche Regeln bei Formulierung von SPSS-Befehlen zu beachten sind², wird erläutert, bevor im anschließenden Kapitel auf Datenbeschreibung und Datenerfassung eingegangen wird.³ Möglichkeiten

1 Der Vorschlag, die Kommandos im Log auflisten zu lassen, kann nur nachdrücklich unterstützt werden, da die Möglichkeit Kommandos und output direkt nacheinander lesen zu können, hilfreich ist bei der Fehlersuche und -bereinigung und bei der Interpretation umfangreicherer Analysen.

2 Hier findet sich der oft gegebene Rat hinsichtlich der Benennung von Variablen nach mnemotechnischen Gesichtspunkten; diese Vorgehensweise ist m. E. nur begrenzt sinnvoll, etwa bei einer überschaubaren Anzahl von Variablen.

Der Hinweis auf die Existenz von Systemvariablen ist selten in SPSS-Einführungen zu finden, ergänzt werden sollte, dass durch eine dauerhafte Speicherung insbesondere von \$casenum (COMPUTE id=\$casenum) eine laufende Identifikationsnummer erstellt werden kann.

3 In diesem Zusammenhang ist der Hinweis auf die standardmäßige rechtsbündige Eingabe der Datenwerte in eine Rohdatendatei zu würdigen, eine abweichende Eingabe führt häufig zu Fehlern.

der Datentransformation, (recode⁴, compute, if, weight, count), der Datenauswahl (select if, sample), SPSS-Datendateien (get, save, display, list, sort cases, split file, export, import) werden beschrieben, bevor im dritten großen Kapitel anhand des ALLBUS2000-Datensatzes Statistikprozeduren vorgestellt werden.

Die Autoren unterscheiden Datenanalysen in explorative Datenanalyse zur Prüfung der Datenqualität oder zur inhaltlichen Exploration von Faktoren oder Clustern und in deskriptive und konfirmative Datenanalyse, um Daten zu beschreiben, zu interpretieren, um Rückschlüsse von einer Stichprobe auf die Grundgesamtheit zu ziehen, Hypothesen zu prüfen oder kausale Beziehungen zwischen Variablen zu untersuchen. Explorative Datenanalysen werden in univariate (descriptives, frequencies⁵) und bivariate (examine, scatterplot)⁶ unterteilt. Als multivariate explorative Analysen werden die Reliabilität von Indices geprüft (reliability), Faktorenanalysen zur Prüfung hinsichtlich der Dimensionalität aber auch zur Datenreduktion durchgeführt (factor⁷), und zur Zusammenfassung einander ähnlicher Untersuchungsobjekte zu homogenen Clustern die in SPSS 11.5 neue Prozedur twostep cluster vorggeführt. Als univariate deskriptive und konfirmative Verfahren werden erneut descriptives und frequencies herangezogen, als bivariate deskriptive und konfirmative Analyseverfahren crosstabs, regression, correlations⁸, nonpar corr, t-test, npar tests m-w, oneway oder npar tests k-w zur Überprüfung der Relationen zwischen zwei Variablen, entweder in Form von Zusammenhängen oder in Form von Unterschieden. Nur kurz angerissen werden multivariate deskriptive und konfirmative Analyseverfahren, da eine ausführliche Behandlung des Themas den Rahmen eines solchen Lehr-

4 Bei der Auflistung auf S. 78, welche Input-Schlüsselwörter für numerische und alphanumerische Variablen bei recode verwendbar sind, fehlt für Zeichenkettenvariablen das Schlüsselwort MISSING. Für das Beispiel zu RECODE auf S. 79 ist ein Fehler in der Syntax zu korrigieren; wenn Wertetikette ergänzt oder modifiziert werden sollen, dann wird dies erreicht durch ADD VALUE LABELS, mit VALUE LABELS würde ein unvollständige Neudefinition vorgenommen. Im Beispiel auf S. 80 werden durch RECODE alter_b andere Altersgruppen definiert als in den Wertetiketten angegeben ist, da sich die in den einzelnen Klammern definierten Gruppen überschneiden. Das Kommando sollte wohl heißen: RECODE alter_b (LO THRU 20=1)(21 THRU 30=2)(31 THRU 40=3)...

5 Zur Exploration der Daten wird eine Grundauszählung aller Variablen vorgeschlagen, für den hier vorliegenden ALLBUS2000-Datensatz sollte die Identifikationsnummer ausgeschlossen werden.

6 Die SPSS-Syntax auf S. 129 entspricht nicht dem beschreibenden Text und der menügesteuerten Vorgehensweise bezüglich der Sortierreihenfolge; hier sollte es heißen: SORT CASES BY sex_b(D).

7 Die SPSS-Syntax auf S. 141 entspricht nicht dem beschreibenden Text und der menügesteuerten Vorgehensweise Syntax in der Zeile beginnend mit /PRINT INITIAL muss ergänzt werden durch EXTRACTION ROTATION.

8 Unsinnige letzte Zeile in der SPSS-Syntax auf S. 191 \index {NONPARR CORR}.

buches übersteigen würde: Genannt werden zur Erstellung von Dreivariablenkontingenztabellen crosstabs und partial corr, das die Korrelation zwischen einer unabhängigen und einer abhängigen Variablen bereinigt von dem Einfluß einer oder mehrerer Kontrollvariablen, also Partialkorrelationen, berechnet.

In den letzten beiden Kapiteln werden einige der in SPSS 11.5 enthaltenen Präsentationsgrafiken⁹ und Präsentationstabellen dargestellt. Balkendiagramme, Kreisdiagramme¹⁰ und Streudiagramme und deren Editiermöglichkeiten werden genauer erläutert¹¹. Leider wird die Möglichkeit SPSS-Grafiken zu exportieren, um sie in anderen Programmen verwenden zu können, nicht demonstriert. Auch auf Einschränkungen hinsichtlich der Weiterverarbeitung der in SPSS erstellten Grafiken in anderen Programmen wird nicht hingewiesen. Wie bei den Präsentationsgrafiken gibt es auch bei den mit SPSS zu erstellenden Präsentationstabellen eine größere Vielfalt von Darstellungsmöglichkeiten, von denen nur einige detaillierter erläutert werden konnten, und dies betrifft überwiegend einfache Tabellen. Wie die sich ergebenden Pivot-Tabellen editiert werden, etwa wie Spalten und Zeilenvariablen vertauscht werden können, wird anschaulich beschrieben. Auch hier fehlt der Hinweis auf die Weiterverarbeitungsmöglichkeiten außerhalb von SPSS, die für Präsentationstabellen besser ausfallen als für -grafiken. Der Anhang enthält neben den SPSS-Kommandos zur Erstellung der Datei, mit der hier gearbeitet wird, die Lösung der Übungsaufgaben und die Information, wie und wo der ALLBUS 2000-Datensatz, die Beispielprogramme und deren Lösungen heruntergeladen werden können¹².

9 Um Diagramme schwarz/weiß und mit unterschiedlichen Füllmustern zu erstellen, muss die Einstellung der Programmoption in der Registerkarte Diagramme "Füllmuster und Linienstil / Muster durchlaufen" ausgewählt werden (vergleiche S. 18ff.).

10 Die SPSS-Syntax auf S. 229 entspricht nicht dem beschreibenden Text, nach dem Diagramm mit Titel und Fußnote versehen sein sollten.

In der menügesteuerten Erstellung des Kreisdiagramms auf S. 231 fehlt der Eingabetext für die zweite Titelzeile "Median", die sich im Pendant der Befehlssyntax und im präsentierten Kreisdiagramm selbst wiederfindet.

11 Da alle hier besprochenen Grafiken mit Titelzeilen und Fußnoten versehen werden können, sollten sie, sofern sie auf gewichteten Daten beruhen, eine Notiz enthalten über die Gewichtung der Daten.

12 Leider ist im Mai 2004 dieser Zugang nicht mehr aktiv gewesen; die Unterlagen und Daten sind derzeit zugänglich unter:

<http://www.rzr.uni-erlangen.de/dienste/software/produkte/allbus2000/>

"Datenanalyse mit SPSS für Windows" von Reinhard Wittenberg und Hans Cramer ist ein Hand- und Lehrbuch, das alles in allem und ungeachtet kleinerer Fehler zum Kauf empfohlen werden kann. Besonders geeignet ist das vorliegende Buch für Studenten, die bereits methodische Vorkenntnisse haben und nach einer fundierten Einführung in Datenerfassung und Datenanalyse mit SPSS suchen.

EVI SCHOLZ

ZUMA-METHODENBERICHTE

In der ZUMA-Reihe „Methodenberichte“ werden Beiträge publiziert, die wichtige Aspekte der Arbeit von ZUMA dokumentieren, aber in der Regel nicht für eine Veröffentlichung in einer Fachzeitschrift geeignet sind, wie z.B. Dokumentationen, Bibliographien, Literaturberichte und Datenbeschreibungen. Die Berichte sind im Internet als PDF-Datei abrufbar (www.gesis.org/publikationen/berichte/zuma_methodenberichte/). Einzelhefte können auch bei folgender Adresse bestellt werden:

Zentrum für Umfragen, Methoden und Analysen
ZUMA-Publikationen
Postfach 12 21 55
68072 Mannheim

Michael Blohm & Roland Schnell:

ALLBUS-Bibliographie (19. Fassung, Stand: November 2003). ZUMA-Methodenbericht 2003/13.

Sabine Klein & Janet Harkness:

ISSP Study Monitoring 2001. Report to the ISSP General Assembly on monitoring work undertaken für the ISSP by ZUMA, Germany. ZUMA-Methodenbericht 2003/14.

Evi Scholz, Janet Harkness & Sabine Klein:

ISSP 2002 Germany. Family and Changing Gender Roles III. ZUMA-Report on the German Study. ZUMA-Methodenbericht 2003/15.

Evi Scholz, Janet Harkness & Sabine Klein:

ISSP 2001 Germany. Social Networks II: Social Relations and Support Systems. ZUMA-Report on the German Study. ZUMA-Methodenbericht 2003/16.

Antje Springer & Georgios Papastefanou:

EVS-Bibliographie (3. Auflage, Stand: November 2003). ZUMA-Methodenbericht 2004/01.

Janet Harkness & Alexander Haarmann:

German General Social Survey 2000. English Translation of the German „ALLBUS“-Questionnaire (Stand: Januar 2004). ZUMA-Methodenbericht 2004/02.

Janet Harkness & Michael Blohm:

German General Social Survey 2002. English Translation of the German „ALLBUS“-Questionnaire. ZUMA-Methodenbericht 2004/03.

Antje Springer & Georgios Papastefanou:

Bibliografie zu ausgewählten Analysen der Luxembourg Income Study (LIS) (2. Auflage, Stand: 12/2003) ZUMA-Methodenbericht 2004/04.

ZUMA „HOW TO-REIHE“

Die ZUMA „How to-Reihe“ ist eine Schriftenreihe, in der ausgewählte Themen aus dem Bereich der sozialwissenschaftlichen Methodik auf unpräzise anwendungsorientierte Weise dargestellt werden. Der besondere Akzent liegt auf der Vermittlung von Know-how, das für den Anwender wichtig und nützlich ist. Gezeigt werden soll vor allem, wie ganz konkret zu verfahren ist. Die Beiträge der Reihe finden Sie unter http://www.gesis.org/Publikationen/Berichte/ZUMA_How_to/.

Ingwer Borg:

Explorative Multidimensionale Skalierung. How-to 1, 2000

Rolf Porst:

Question Wording – Zur Formulierung von Fragebogen-Fragen. How-to 2, 2000

Michael Braun:

Evaluation der Äquivalenz eines gemeinsamen Satzes an Indikatoren in der interkulturell vergleichenden Sozialforschung. How-to 3, 2000

Jürgen H.P. Hoffmeyer-Zlotnik:

Regionalisierung von Umfragedaten. Eine kleine Handlungsanweisung. How-to 4, 2000

Michael Häder:

Die Expertenauswahl bei Delphi-Befragungen. How-to 5, 2000

Sabine Häder:

Telefonstichproben. How-to 6, 2000

Jürgen H.P. Hoffmeyer-Zlotnik:

Wohnquartiersbeschreibung: Ein Instrument zur Regionalisierung von Nachbarschaften. How-to 7, 2001

Cornelia Züll & Peter Ph. Mohler:

Computerunterstützte Inhaltsanalyse: Codierung und Analyse von Antworten auf offene Fragen. How-to 8, 2001

Rolf Porst:

Wie man die Rücklaufquote bei postalischen Befragungen erhöht. How-to 9, 2001

Michael Wiedenbeck & Cornelia Züll:

Klassifikation mit Clusteranalyse: Grundlegende Techniken hierarchischer und k-means-Verfahren. How-to 10, 2001

Peter Prüfer & Angelika Stiegler:

Die Durchführung standardisierter Interviews: Ein Leitfaden. How-to 11, 2002

Beatrice Rammstedt:

Zur Bestimmung der Güte von Multi-Item-Skalen: Eine Einführung. How-to 12, 2004

ZUMA-WORKSHOPS

Zweites Halbjahr 2004

Thema	Datum
Using (Multidimensional) Rasch and other IRT Models	2. September
Einführung in die Analyse linearer Strukturgleichungsmodelle mit LISREL 8.54	23. – 24. September
Empirische Elitenforschung	27. – 28. September
Simulation for the Social Sciences	27. September – 1. Oktober
Der Mikrozensus als Datenquelle für die empirische Sozial- und Wirtschaftsforschung: Arbeiten mit den Daten	29. – 30. September
Stichproben in Theorie und Praxis	5. Oktober
Einführung in die Ereignisanalyse von Verbraucherpaneldaten	14. Oktober
Interkultureller Vergleich von Einstellungsvariablen	19. – 20. Oktober
Einführung in die computerunterstützte Inhaltsanalyse (cui)	9. – 10. November
Demographische Standards: Empirische Grundlagen der Sozialstrukturanalyse	11. – 12. November
Introduction to Working with Cross-National Surveys: The Eurobarometer	16. November
Analyse linearer Strukturgleichungsmodelle mit LISREL 8.54 für Fortgeschrittene	18. – 19. November
Fragebogen: Entwicklung und empirische Qualitätsüberprüfung	23. – 24. November
Praktische Durchführung von Telefonumfragen	26. November
Subjektives Wohlbefinden und subjektive Indikatoren: Konzepte, Daten und Analysen	2. – 3. Dezember

Detaillierte Hinweise zu den oben aufgeführten Veranstaltungen erhalten Sie unter:
<http://www.gesis.org/Veranstaltungen/ZUMA/Workshops/index.htm>
 oder beim ZUMA-Tagungssekretariat (workshop@zuma-mannheim.de, Tel.: 0621/1246221).

NACHRICHTEN AUS DEM INSTITUT

Besondere Anerkennung für ihre Arbeiten bei ZUMA haben CAROLINE CRAMER und JÜRGEN HOFFMEYER-ZLOTNIK erfahren.

CAROLINE CRAMER, in der Abteilung Sozialindikatoren tätig, wird im Sommersemester 2004 und im Wintersemester 2004/05 an der Ludwig-Maximilians-Universität München eine Professur für Stadtgeographie vertreten.

JÜRGEN HOFFMEYER-ZLOTNIK vertritt im Sommersemester 2004 an der Universität Duisburg-Essen, Campus Essen, eine Professur für Empirische Sozialforschung.

Für ZUMA eigentlich nicht erfreulich, aber ebenfalls als Anerkennung ihrer wissenschaftlichen Arbeit zu sehen: ANNETTE KOHLMANN, bisher Mitarbeiterin im German Microdata Lab, hat zum 1. April 2004 die Leitung des Forschungsdatenzentrums der Bundesagentur für Arbeit am Institut für Arbeitsmarkt- und Berufsforschung (IAB) in Nürnberg übernommen.

Last but not least: LARS KACZMIREK, seit Mai 2003 im Projekt Web Survey Methodology Site (WebSM) bei ZUMA tätig, erhielt im Februar 2004 den Otto-Selz-Preis 2004 für die beste Diplomarbeit des Fachbereiches Psychologie an der Sozialwissenschaftlichen Fakultät der Universität Mannheim.

ZUMA verlassen haben NADIA GRANATO und Immo WITTIG, beide bisher im German Microdata Lab tätig. An der Telefonzentrale des Hauses hat es einen Wechsel gegeben: Anstelle von HANNA FISCHER, die in den Ruhestand gegangen ist, empfängt Sie seit dem 1. April 2004 MONIKA LÖW.

DURCHWAHL-RUFNUMMERN (STAND: MAI 2004)

Sie erreichen die Mitarbeiter von ZUMA unter der Nummer 0621-1246-(Durchwahlnummer); die Zentrale unter 1246-0. Sie ist von Montag bis Donnerstag von 8.30 bis 17.00 und freitags von 8.30 bis 15.30 besetzt. Die mit (S) bezeichneten Mitarbeiterinnen nehmen Sekretariatsaufgaben wahr.

DIREKTION

<i>Prof. Dr. Peter Ph. Mohler (Direktor)</i>	173
Carol Cassidy (Stellv. Direktorin)	146
Margit Bäck (S)	172
Elisabeth Bähr (S)	172
Maria Kreppe-Aygün	184

INTERNE INFRASTRUKTUR**Verwaltung**

Dipl.-Kfm. Jost Henze	161
-----------------------	-----

Information & Kommunikation

Dipl.-Soz. Kerstin Hollerbach	174
-------------------------------	-----

EDV-Infrastruktur

Carol Cassidy	146
---------------	-----

Datenbanken

Joachim Wackerow	262
------------------	-----

WISSENSVERMITTLUNG & BERATUNG**Wissenschaftlicher Leiter**

Prof. Dr. Ingwer Borg	151
-----------------------	-----

Projektberater

Dr. Wolfgang Bandilla	136
Dr. Michael Braun	176
PD Dr. Jürgen H.P. Hoffmeyer-Zlotnik	175
Dipl. Soz. Rolf Porst	228
Dr. Beatrice Rammstedt	155
Christa v. Briel (S)	231
Dagmar Haas (S)	152
Patricia Lüder (S)	221

Pretesting

Dipl.-Psych. Peter Prüfer	227
Margrit Rexroth, M.A.	230

Textanalyse, Vercodung

Alfons J. Geis, M.A.	222
Patricia Lüder (S)	221

Statistik

PD Dr. Siegfried Gabler	281
Dr. Sabine Häder	282
Dipl.-Math. Michael Wiedenbeck	279

Telefonumfragen

Dipl.-Soz. Michael Schneid	209
Dipl.-Soz. Angelika Stiegler	208

Online-Umfragen

Dipl.-Sozialw. Wolfgang Neubarth	205
----------------------------------	-----

Elektronische Handbücher ZIS/EHES

Dr. Angelika Glöckner-Rist	171
----------------------------	-----

Computerunterstützte Textanalyse, Textpack, NSD-stat

Cornelia Züll	147
Juliane Landmann	144

DAUERBEOBACHTUNG**Wissenschaftliche Leiterin**

Prof. Dr. Ursula Hoffmann-Lange	247
---------------------------------	-----

ALLBUS

<i>Dipl.-Soz. Achim Koch</i>	280
Dipl. Soz. Michael Blohm	276
Dipl. Soz. Alexander Haarmann	286
Dipl.-Soz. Martina Wasmer	273
Julia Khorshed (S)	274

International Social Survey Programme (ISSP)

<i>Dr. Janet Harkness</i>	284
Sabine Klein	272
Evi Scholz	283

German Micro Data Lab

Prof. Dr. Ursula Hoffmann-Lange	247
<i>Dipl.-Soz. Bernhard Schimpl-Neimanns</i>	263
Dipl.-Soz. Jeanette Bohr	261
Dipl.-Sozialwiss. Helga Christians	266
Dipl. Soz. Nadia Granato	264
Dr. Paul Lüttinger	268
<i>Dr. Georgios Papastefanou</i>	278
Dr. Heike Wirth	269
Irene Fischer (S)	265

Soziale Indikatoren

<i>Dr. Heinz-Herbert Noll</i>	241
Dipl.-Soz. Regina Berger-Schmitt	248
Dr. Caroline Kramer	244
Dr. Stefan Weick	245
Margit Bäck (S)	242

DRITTMITTELPROJEKTE

Nina Rother (PIONEUR)	285
Lars Kaczmirek (WebSM)	206

HINWEISE FÜR UNSERE AUTOREN

Die ZUMA-Nachrichten veröffentlichen - neben Nachrichten des Instituts - Artikel, die von Interesse für die empirische Sozialforschung, insbesondere die Umfrageforschung, sind. Alle Artikel müssen einen methodischen Fokus haben oder zumindest methodische Aspekte stark betonen. Das Spektrum möglicher Artikel ist breit: Es reicht von Grundlagenforschung über angewandte Papiere bis zu Arbeiten, die einen praktisch-operativen Charakter haben. Die Artikel in den ZUMA-Nachrichten sollen für eine breite Leserschaft von Wissenschaftlern und Praktikern im Bereich der empirischen Sozialforschung verständlich sein.

Alle Beiträge, die zur Veröffentlichung in den ZN eingereicht werden, werden von mindestens zwei unabhängigen Gutachtern blind begutachtet.

Die folgenden Regeln sind bei der Abfassung von Manuskripten zu beachten:

1. Manuskripte müssen per email (porst@zuma-mannheim.de) eingereicht werden. Der Umfang der Manuskripte (einseitig und 1,5-zeilig beschrieben, Punktgröße 11) soll alles in allem nicht mehr als 30 Seiten betragen.
2. Den Beiträgen sind Abstracts in Deutsch und Englisch (jeweils ca. 15 Zeilen) voranzustellen.
3. Beiträge sind mit dem Dezimalklassifikationssystem zu untergliedern (1 – 2 – 2.1 – 2.2 – 3 usw.). Die Gliederungstiefe geht dabei höchstens auf *eine* Stelle nach dem Punkt.
4. Tabellen enthalten Tabellenummer und Titel im Tabellenkopf, Abbildungen werden analog behandelt.
5. Grafiken sind mittels gängiger Grafiksoftware zu erstellen. Ist eine spezielle Grafiksoftware erforderlich, übernimmt der Autor/die Autorin die endgültige Formatierung der Grafiken in eigener Regie.
6. Bei der Erstellung von Tabellen und Grafiken ist zu berücksichtigen, dass die ZUMA-Nachrichten DIN A5-Format haben und der Satzspiegel 11,5 cm (Breite) x 16 cm (Höhe) beträgt.
7. Anmerkungen und Fußnoten sind mit der Fußnotenfunktion des Schreibprogrammes (im Normalfalle WORD) zu erstellen; bitte nicht gesondert formatieren. Fußnoten sind nur für inhaltliche Kommentare vorzusehen, nicht für bibliographische Hinweise.
8. Literaturhinweise im Text sind nach den folgenden Mustern aufzuführen: Müller (2002) / (Müller, 2002) / Müller (2002:75) / (vgl. Müller, 2002:75) / (Müller, 2002; Mayer & Müller, 2003).
9. Das Literaturverzeichnis ist gemäß den Richtlinien der American Psychological Association (APA) zu gestalten. Beispiele:
Zeitschriftenbeiträge:
Salzgeber, J. & Stadler, M. (1997). Programm zur Behandlung von Sexualstraftätern. *Zeitschrift für Rechtspsychologie*, 4, 139-141.
Buchveröffentlichungen:
Nicol, A. A. & Pexman, P. M. (1999). *Presenting your findings: A practical guide for creating tables*. Washington, DC: American Psychological Association.
Beiträge in Büchern:
O'Neil, J. M. & Egan, J. (1992). Men's and women's gender role journeys: Metaphor for healing, transition, and transformation. In B. R. Wainrib (Hrsg.), *Gender issues across the life cycle* (S. 107-123). New York: Springer.
10. Die Beiträge sind unter Wahrung der gültigen Rechtschreibregelungen (neue Rechtschreibung) zu erstellen.