GESIS

ZUMA

NACHRICHTEN

51

November 2002

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 - 268
E-Mail luettinger@zuma-mannheim.de
Internet www.gesis.org/zuma ZUMA
www.gesis.org/ GESIS

Redaktion: Dr. Paul Lüttinger

ISSN 0941-1670 26. Jahrgang

Die ZUMA-Nachrichten informieren über Forschungsarbeiten und Dienstleistungen des Instituts. Beiträge externer Autoren werden dann aufgenommen, wenn sie aus Kooperationsprojekten mit ZUMA resultieren oder von besonderem wissenschaftlichen Interesse für die Arbeiten im Hause und die Empirische Sozialforschung sind. Alle Beiträge unterliegen einem internen förmlichen Begutachtungsverfahren durch jeweils zwei Gutachter. Der Herausgeber entscheidet auf der Grundlage der Gutachten über die Annahme eines Beitrages. Hinweise zur Manuskriptgestaltung sind auf der letzten Seite des Heftes.

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: http://www.gesis.org/publikationen/zeitschriften/zuma_nachrichten/.

Druck: Verlag Pfälzische Post GmbH, Neustadt/Weinstraße.
Gedruckt auf chlorfrei gebleichtem Papier.

© **ZUMA, Mannheim, 2002**

In eigener Sache 5

In Memoriam Dr. Peter Petersen 7

FORSCHUNGSBERICHTE

20 Jahre Feldarbeit im ALLBUS: Ein Blick in die Blackbox.
Achim Koch 9

Wer nimmt teil an Panel-Befragungen? Untersuchung über die Bedingungen der erfolgreichen Kontaktierung für sozialwissenschaftliche Untersuchungen.
Tatjana Mika 38

Auswirkungen des Erhebungsverfahrens bei Jugendbefragungen zu „heiklen Themen“ - schulbasierte schriftliche Befragung und haushaltsbasierte mündliche Befragung im Vergleich.
Dietrich Oberwittler & Thomas Naplava 49

Generierung eins Proxys zum Job-Anforderungsniveau aus den Informationen zu ausgeübtem Beruf und beruflicher Stellung. Ein neues Tool für die deutsche Überqualifikations-Forschung.
Matthias Pollmann-Schult & Felix Büchel 78

MITTEILUNGEN

Von der Schwierigkeit, Rechtsfragen sozialwissenschaftlich zu untersuchen. Replik zu Häder/Klein, ZUMA-Nachrichten 50, S. 86 – 111.
Arthur Hartmann 94

Anmerkungen zu Arthur Hartmann: Von der Schwierigkeit, Rechtsfragen sozialwissenschaftlich zu untersuchen. Replik zu Häder/Klein.
Michael Häder & Sabine Klein 104

Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS): Planung zukünftiger Befragungsschwerpunkte.
Hans-Jürgen Andreß 117

Edition von Lebensverlaufsdaten: zur Relevanz einer systematischen Einzelfallbearbeitung bei standardisierten Befragungen.
Steffen Hillmert 120

Scientific-Use-File des Mikrozensus 1998 für die Wissenschaft zugänglich 141

LIS-Sommerworkshop 2003: Internationale Vergleiche der Einkommensungleichheit 143

BUCHBESPRECHUNG

THOMAS PETERSEN: Das Feldexperiment in der Umfrageforschung 145

PUBLIKATIONEN

ZUMA-Methodenberichte 155
Buchhinweise 156

VERANSTALTUNGEN 2003

Workshop: Mitarbeiterbefragungen, 18. – 19. März 162
Workshop: Sozio-demographische Variablen in der Umfrageforschung, 25. – 26. März ... 163
Workshop: Einführung in die Clusteranalyse, 1. – 2. April 163
Workshop: Einführung in Datenmanagement und Datenrecodierung, 16. – 17. Juni 164
Workshop: Evaluation von Studiengängen – Methodische Aspekte, 24. – 25. Juni 165
Workshop: Einführung in die Analyse linearer Strukturgleichungsmodelle mit LISREL 8.52, 26. – 27. Juni 165
Workshop: Hermeneutic Procedures to Improve Questionnaire Research, 2. – 3. Juli 166
Workshop: ‚Political Polls': Politische Umfragen in den Massenmedien, 10. – 11. Juli ... 167
Workshop: Survey Design for the Multinational Context, 4. – 5. September 167
Workshop: Simulation for the Social Sciences, 8. – 12. September 168
Workshop: Analyse linearer Strukturgleichungsmodelle mit LISREL 8.52 für Fortgeschrittene, 25. – 26. September 169
Workshop: ADM-Stichprobensysteme in der Umfragepraxis: CATI und face-to-face, 30. Sept. – 1. Oktober 169
3. Nutzerkonferenz: Forschung mit dem Mikrozensus, 9. – 10. Oktober 171

DURCHWAHL-RUFNUMMERN

IN EIGENER SACHE

Über viele Details der Feldarbeit bei persönlich-mündlichen (face-to-face) Bevölkerungsbefragungen ist bislang nur wenig bekannt. ACHIM KOCH beleuchtet in seinem Beitrag „20 Jahre Feldarbeit im ALLBUS: Ein Blick in die Blackbox“ einige ausgewählte Facetten anhand der zwölf Erhebungen des ALLBUS (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften) seit 1980. Betrachtet werden u.a. Anzahl und demographische Zusammensetzung der eingesetzten Interviewer, Wochentage und Tageszeiten der Interviewdurchführung sowie Informationen zur Feldzeit und zu den Fallpreisen in den einzelnen Erhebungen.

TATJANA MIKA berichtet im Beitrag „Wer nimmt Teil an Panel-Befragungen?“ über den Einfluss verschiedener Faktoren auf die Bereitschaft, sich für sozialwissenschaftliche Untersuchungen befragen zu lassen.

Schriftliche Befragungen im Klassenverband sind heute in der empirischen Jugendforschung weit verbreitet, ohne dass die methodischen Aspekte dieser Erhebungsform als ausreichend untersucht gelten können. *DIETRICH OBERWITTLER & THOMAS NAPLAVA* beschäftigen sich mit den Auswirkungen des Erhebungsverfahrens bei Jugendbefragungen und vergleichen eine schulbasierte schriftliche Befragung mit einer haushaltsbasierten mündlichen Befragung zum Thema selbstberichtete Delinquenz.

In der Arbeitsmarktforschung wird als Fehlqualifikation eine fehlende Übereinstimmung zwischen erworbener beruflicher Qualifikation und den Anforderungen der ausgeübten beruflichen Tätigkeit bezeichnet. Da in den verfügbaren Mikrodaten häufig keine hinreichend genauen Angaben zu den Anforderungen der ausgeübten beruflichen Tätigkeit vorhanden sich behelfen sich viele Forscher dadurch, dass sie die Information zur beruflichen Stellung als Proxy des Job-Anforderungsniveaus verwenden. MATTHIAS POLLMANN-SCHULT & FELIX BÜCHEL stellen ein neues Verfahren zur Generierung eines Proxys des Job-Anforderungsniveau vor, das auf den Informationen zum ausgeübten Beruf und zur beruflichen Stellung beruht.

RÜDIGER SCHMITT-BECK hat einen Ruf auf eine C3 Professur an die Universität Duisburg erhalten.

Peter Ph. Mohler

IN MEMORIAM DR. PETER PETERSEN

Am 11. Oktober 2002 ist Dr. Peter Petersen in Bonn im Alter von 73 Jahren gestorben. Damit endet eine ganz besondere Beziehung.

Zum Zeitpunkt der Gründung von ZUMA am 1. Januar, 1974 war Dr. Petersen Fachreferent der DFG für Politische Wissenschaft. Rudolf Wildenmann, der Gründer von ZUMA, war mit Dr. Petersen unter anderem durch Wildenmanns Mitgliedschaft im Senat der DFG in engeren Kontakt gekommen. Beide waren sich einig in der Einschätzung, dass nicht nur die Natur-, sondern auch die Sozialwissenschaften einer Forschungsinfrastruktur bedürften, ein damals revolutionärer Gedanke. Nicht zuletzt durch diesen Kontakt fühlte sich Rudolf Wildenmann ermutigt, seine aus den Erfahrungen mit der Kölner Wahlstudie von 1961 gewonnene Einschätzung konsequent zu entwickeln, zu einer solchen Infrastruktur mit einem akademischen Institut zur universitären Umfrageforschung beizutragen.

Im Dialog mit der DFG, deren damaliger Präsident Prof. Dr. Julius Speer solchen Absichten aufgeschlossen gegenüberstand, und zahlreichen Fachwissenschaftlern, darunter M. Rainer Lepsius, Burkhart Lutz, Werner Mangold und Erwin K. Scheuch, entstand so der Plan zur Gründung des Zentrums für Umfragen, Methoden und Analysen in Mannheim als Hilfseinrichtung der Forschung der DFG. Dass dieses Konzept dann auch in die Tat umgesetzt werden konnte, war nicht zuletzt der Unterstützung von Dr. Petersen zu danken.

Mit dem DFG-Präsidentenwechsel 1974 entstand eine neue, schwierigere Situation für ZUMA, das ja just zu diesem Zeitpunkt seine Arbeit aufnahm. Blicke zurück tendieren gelegentlich dazu, Situationen und Entwicklungen falsch einzuschätzen. Diejenigen, die in jener Zeit für ZUMA Aufbauarbeit geleistet haben, sind jedoch übereinstimmend der Auffassung, dass die schützende und führende Hand von Dr. Petersen mit eine entscheidende Voraussetzung dafür war, dass ZUMA diese problematische Anfangsphase unbeschadet überstanden hat und die Evaluation durch den Wissenschaftsrat 1981 diesen Aufbauerfolg zertifiziert hat.

Als erster Direktor des Instituts von 1974 bis 1980 kann ich mich an zahlreiche Situationen erinnern, in denen es der Rat und die Erfahrung von Dr. Petersen waren, die dazu beigetragen haben, Fehler zu vermeiden und die im Rahmen der DFG-Finanzierung verfügbaren Möglichkeiten optimal zu nutzen.

Nach dem Eintritt von ZUMA in die GESIS als Institut der Blauen Liste war die institutionelle Verbindung zur DFG naturgemäß beendet, nicht aber die persönliche zu Dr. Petersen, inzwischen Leiter der Verfahrensgruppe der DFG. Als Mitglied der Senatsausschüsse ab 1983 für die Sonderforschungsbereiche, die Graduiertenkollegs sowie Senat und Hauptausschuss habe ich bei meinen Besuchen in der DFG-Geschäftsstelle selten die Gelegenheit zu einem persönlichen Gespräch mit ihm ausgelassen, und ZUMA war dabei stets auch ein Thema. Und noch 1999, bereits als Pensionär, war er in Mannheim unter den Gästen der ZUMA-Feier zum 25jährigen Bestehen des Instituts.

Im Herbst 1994 ist Dr. Petersen aus dem aktiven Dienst bei der DFG ausgeschieden. Nun ist er überraschend, im Alter von nur 73 Jahren, verstorben. Diejenigen bei ZUMA, die ihn persönlich gekannt haben, wissen, was ich zum Ausdruck bringen will, wenn ich sage: Peter Petersen hat sich um ZUMA verdient gemacht. ZUMA trauert mit seiner Frau und seinen Kindern um diesen Verlust.

Max Kaase, Bremen, im Oktober 2002

20 JAHRE FELDARBEIT IM ALLBUS: EIN BLICK IN DIE BLACKBOX[1]

ACHIM KOCH

If you love surveys or sausages, you should not watch either being made.
(J. Kochevar)

Über die Details der Datenerhebung bei persönlich-mündlichen (face-to-face) Bevölkerungsbefragungen ist nur wenig bekannt. Mit dem vorliegenden Beitrag sollen schlaglichtartig einige ausgewählte Facetten der Feldarbeit beleuchtet werden. Datengrundlage bilden die zwölf Erhebungen, die für die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) zwischen 1980 und 2000 durchgeführt wurden. Betrachtet werden u.a. Anzahl und demographische Zusammensetzung der eingesetzten Interviewer, Wochentage und Tageszeiten der Interviewdurchführung sowie Informationen zur Feldzeit und zu den Fallpreisen in den einzelnen Erhebungen. Vor dem Hintergrund der Problematik der Ausschöpfungsberechnung wird zum Schluß auf die Frage des Zusammenhangs von Qualität, Zeitbedarf und Preis eingegangen.

The level of knowledge about the details of fieldwork in face-to-face interview surveys of the general population is rather low. The present paper focuses on certain particular aspects of fieldwork. The database consists of the twelve surveys of the German General Social Survey (ALLBUS), fielded between 1980 and 2000. We investigate the number and demographic structure of the interviewers, the days and times of interviewing, as well as both the length of the fielding period and the costs of fielding. The relationship between quality, fielding time, and costs is discussed against the background of the problems involved in determining response rates in a reliable manner.

1 Überarbeitete Version eines Vortrages auf der Frühjahrstagung der Sektion Methoden der DGS am 30./31. März 2001 in Konstanz.

1. Einleitung

Die Datenerhebung bei persönlich-mündlichen (face-to-face) Bevölkerungsbefragungen erfolgt in Deutschland ganz überwiegend durch privatwirtschaftlich verfaßte Umfrageinstitute. Diese arbeiten mit einem Stab von Interviewern zusammen, die in der Regel als freiberufliche Mitarbeiter für die Institute tätig sind. Die Aufgaben der Interviewer bei der Feldarbeit umfassen die Mitwirkung an der Stichprobenbildung, die Kontaktierung der Zielpersonen und deren Motivierung zur Teilnahme an der Studie und schließlich die Durchführung der Befragung entsprechend den Regeln des standardisierten Interviews. Die Umfrageinstitute setzen u.a. durch ihre Verfahren der Interviewerschulung und –kontrolle, durch die verwendeten Honorierungsmodelle sowie weitere studienbezogene Vorgaben, z.B. hinsichlich der Zahl und zeitlichen Terminierung von Kontaktversuchen, die Rahmenbedingungen für die Tätigkeit der Interviewer.

Sowohl über das Institutshandeln als auch über das tatsächliche Verhalten der Interviewer, über typische Probleme der Feldarbeit und deren Lösung unter Praxisbedingungen ist indes nur wenig bekannt. Die Gründe hierfür sind vielfältiger Natur: die Wettbewerbssituation der Umfrageinstitute und deren Angst vor einem möglichen Imageschaden durch die Offenlegung eventuell vorhandener Defizite; fehlende Informationen über das faktische Verhalten der Interviewer auch in den Instituten, weil dieses bei persönlichen – im Gegensatz zu zentral im Studio durchgeführten telefonischen – Umfragen schwieriger zu standardisieren und kontrollieren ist; mangelndes Interesse der Auftraggeber an Fragen der Datenerhebung, etc.

Auch mit dem vorliegenden Beitrag kann lediglich eine erste Beschreibung geleistet werden: Schlaglichtartig werden einige ausgewählte Facetten der Feldarbeit beleuchtet - wohl wissend, dass weite Bereiche weiterhin im Dunkeln verbleiben. Zugrunde liegen Informationen aus den zwölf Erhebungen, die für die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) zwischen 1980 und 2000 durchgeführt wurden. Um die Aussagen auf eine breitere Basis zu stellen, beschränken wir uns im folgenden auf Aspekte, über die jeweils für mehrere – im besten Fall für alle zwölf – Erhebungen Informationen vorliegen.[2] Die gewählte Vorgehensweise erlaubt es, an einigen Stellen Unterschiede und Veränderungen in der Feldarbeit – zwischen verschiedenen Umfrageinstituten und Stichprobenverfahren oder auch über die Zeit – zu identifizieren. Ursachen

---

2 Den Auswertungen liegen einige wenige Merkmale zugrunde. Dies sind eine Kennung der primären Stichprobeneinheit (im folgenden auch als „Sample Point“ bzw. „Klumpen“ bezeichnet), die Interviewer-ID, Informationen zu Geschlecht, Alter und Bildung der eingesetzten Interviewer sowie Angaben zu Datum und Uhrzeit des Interviews. Es wäre wünschenswert, wenn zumindest diese rudimentären Informationen standardmäßig in sozialwissenschaftlichen Umfragen erfaßt würden.

und Gründe als auch mögliche Konsequenzen (z.B. was die Qualität der erhobenen Daten angeht) können nur kursorisch behandelt werden.

In Abschnitt zwei werden zunächst die Durchführungsmodalitäten der ALLBUS-Umfrage, speziell die verwendeten Stichprobenverfahren, skizziert. Daran schließen sich in Abschnitt drei Informationen zur Zahl der in den einzelnen Erhebungen eingesetzten Interviewer und zur Zahl realisierter Interviews pro Interviewer an. In Abschnitt vier wird die Zusammensetzung der eingesetzten Interviewer nach den Merkmalen Geschlecht, Alter und Bildung beleuchtet und analysiert, ob zwischen der Struktur des Interviewerstabes und dem demographischen Profil der realisierten ALLBUS-Stichproben ein Zusammenhang besteht. In Abschnitt fünf wird erläutert, an welchen Wochentagen und zu welchen Tageszeiten die ALLBUS-Interviews durchgeführt wurden. In Abschnitt sechs wird die Feldzeit der ALLBUS-Umfragen beschrieben. Eine Betrachtung auf Ebene der primären Stichprobeneinheiten („Klumpen“) bietet dabei einen detaillierteren Einblick in den zeitlichen Verlauf der Feldarbeit. In Abschnitt sieben wird anhand der durchschnittlichen Fallpreise ein Blick auf die Kosten der Umfragen geworfen. Vor dem Hintergrund der Problematik der Ausschöpfungsberechnung wird zum Schluß in Abschnitt acht auf die Frage des Zusammenhangs von Qualität, Zeitbedarf und Preis eingegangen.

2. Durchführungsmodalitäten der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften

In der ALLBUS-Umfrage wird seit 1980 alle zwei Jahre eine Zufallsstichprobe der Bevölkerung in der Bundesrepublik Deutschland mit einem teils konstanten, teils variablen Fragenprogramm zu sozialwissenschaftlich relevanten Themenstellungen befragt.[3] Zwischen 1980 und 1990 umfaßte die ALLBUS-Stichprobe jeweils ungefähr 3.000 Personen aus der Grundgesamtheit der deutschen Bevölkerung ab 18 Jahren in Privathaushalten der alten Bundesrepublik (vgl. Tabelle 1). 1991 wurde aus Anlaß der deutschen Vereinigung

3 Die Konzeption der Umfrage erfolgt - in Abstimmung mit einem wissenschaftlichen Beirat - durch die Abteilung ALLBUS am Zentrum für Umfragen, Methoden und Analysen (ZUMA) in Mannheim. Zur Verbesserung der Infrastruktur in den Sozialwissenschaften stehen die erhobenen Daten sofort nach ihrer benutzergerechten Aufbereitung und Dokumentation im Zentralarchiv für empirische Sozialforschung (ZA) in Köln interessierten Forschern und Institutionen für Sekundäranalysen zur Verfügung. Der ALLBUS gehört zu einem wachsenden Netz von regelmäßig durchgeführten, unabhängigen nationalen Bevölkerungsumfragen, zu dem auch Studien wie der General Social Survey in den USA, die British Social Attitudes-Erhebungsreihe oder die laufenden Surveys des niederländischen Sociaal en Cultureel Planbureau gehören. Finanziert wird der ALLBUS vom Bund und den Ländern im Rahmen der Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen (GESIS).

zusätzlich eine Umfrage außerhalb des zweijährigen Turnus durchgeführt, bei der erstmals auch Bürger der neuen Bundesländer sowie deutschsprechende Ausländer in die Stichprobe aufgenommen wurden. Die Stichproben der gesamtdeutschen Erhebungen zwischen 1991 und 2000 waren mit circa 3.000 bis 3.500 Personen nicht oder nur geringfügig größer als in den 80er Jahren; Ostdeutschland wurde durch ein „oversampling" in allen Studien bewußt überrepräsentiert.

Tabelle 1: Die ALLBUS-Erhebungen 1980 bis 2000 im Überblick

	1980	1982	1984	1986	1988	1990
Grundgesamtheit	Personen ab 18 Jahren mit deutscher Staatsangehörigkeit in Privathaushalten in Westdeutschland (incl. West-Berlin)					
Stichprobenverfahren*	ADM AR	ADM AR	ADM RR	ADM RR	ADM RR	ADM AR
Nettofallzahl	2955 West	2991 West	3004 West	3095 West	3052 West	3051 West
Befragungsform	PAPI	PAPI	PAPI	PAPI	PAPI	PAPI
Erhebungsinstitut	GETAS	GETAS	GETAS	Infratest	GETAS	Infas
	1991	**1992**	**1994**	**1996**	**1998**	**2000**
Grundgesamtheit	Personen ab 18 Jahren (Deutsche und Ausländer) in Privathaushalten in West- und Ostdeutschland					
Stichprobenverfahren*	ADM SR	ADM SR	Register	Register	ADM AR	Register
Nettofallzahl	1514 West 1544 Ost	2400 West 1148 Ost	2342 West 1108 Ost	2402 West 1116 Ost	2212 West 1022 Ost	2036 West 1102 Ost
Befragungsform	PAPI	PAPI	PAPI	PAPI	PAPI	CAPI
Erhebungsinstitut	Infratest	Infratest	Infratest	Infratest	GETAS	Infratest

* AR = Adress-Random, RR = Random-Route, SR = Standard-Random

Ingesamt wurden in den zwölf ALLBUS-Umfragen zwischen 1980 und 2000 mehr als 38.000 Interviews realisiert. Alle Erhebungen fanden als persönlich-mündliche Befragung statt, bis 1998 als Paper and Pencil Interview (PAPI), im Jahr 2000 zum ersten Mal als Computer Assisted Personal Interview (CAPI). Mit der Feldarbeit waren drei Umfrageinstitute betraut: Infratest Burke (heute NFO Infratest), GFM-GETAS (heute IPSOS Deutschland) und Infas. Sämtliche Umfragen wurden auf der Basis von Zufallsstichproben, entweder in Form des ADM-Designs (bzw. einem äquivalenten Verfahren) oder als Stichproben aus Einwohnermelderegistern, realisiert.

In der Mehrzahl der zwölf ALLBUS-Erhebungen kam das dreistufige Design des Arbeitskreises Deutscher Markt- und Sozialforschungsinstitute (ADM) zum Einsatz, das in

Deutschland das Standard-Stichprobenverfahren bei persönlich-mündlichen Bevölkerungsumfragen darstellt (vgl. Arbeitsgemeinschaft ADM-Stichproben und Bureau Wendt 1994, Behrens/Löffler 1999, Koch 1997). Dieses Verfahren sieht eine dreistufige geschichtete Zufallsauswahl mit den Auswahlstufen ‚Wahlbezirke' – ‚Haushalte' – ‚Personen' vor. In der ersten Stufe wird eine Stichprobe von Wahlbezirken gezogen. In der zweiten und dritten Auswahlstufe ermitteln die Interviewer in den ausgewählten Wahlbezirken nach einem Zufallsverfahren die zu befragenden Haushalte und Zielpersonen. Da für einen Teil der folgenden Ausführungen insbesondere das Vorgehen in der zweiten und dritten Auswahlstufe von Relevanz ist, soll dieses etwas näher erläutert werden.

Für die Wahlbezirke sind im ADM-Design zwar die geographischen Grenzen bekannt, nicht jedoch die Adressen der Haushalte, die sie umfassen. Zur Bestimmung der Zielhaushalte wird deshalb den Interviewern anhand eines zufällig ausgewählten Startpunktes und Fortsetzungsregeln ein Begehungsweg vorgegeben. Entlang dieses Weges haben die Interviewer nach bestimmten Regeln eine festgelegte Zahl von Haushalten („Klingelschilder") aufzulisten.[4] In der Praxis haben sich für die Ermittlung der Auswahleinheiten der zweiten Stufe drei Verfahrensvarianten herausgebildet, die sich insbesondere darin unterscheiden, wieviel Entscheidungsspielraum den Interviewern eingeräumt wird bzw. wieviel Kontrolle von Institutsseite möglich ist. Beim restriktivsten Verfahren, dem sogenannten ‚Adress-Random', erfolgt die Auflistung der Haushaltsadressen durch die Interviewer zeitlich vor der Feldarbeit in einem getrennten Arbeitsschritt. Aus den aufgenommenen Adressen wird im Umfrageinstitut eine Substichprobe von Adressen für die eigentliche Befragung zufällig ausgewählt und den Interviewern zur Bearbeitung übergeben. Verbreiteter als diese Variante sind jedoch Verfahren, bei denen die Haushaltsauflistung in einem Zuge mit der Realisierung der Interviews erfolgt. Beim sogenannten ‚Random-Route' wird den Interviewern eine Bruttozahl von aufzulistenden Adressen vorgegeben, aus denen sie möglichst viele Interviews realisieren sollen. Beim weicheren ‚Standard-Random' wird von den Interviewern lediglich die Realisierung einer bestimmten Nettofallzahl von Interviews gefordert. Das Brutto von Adressen, das dafür angelaufen werden darf, wird entweder nicht begrenzt oder großzügig bemessen (z.B. indem nur fünf Interviews aus maximal 15 Adressen realisiert werden sollen).

Auch die Bestimmung der Befragungsperson im Haushalt – die dritte Auswahlstufe – ist im ADM-Design Aufgabe der Interviewer. Hierzu müssen sie zunächst alle Haushaltsmitglieder, die zur Grundgesamtheit der Studie zählen, in einer festgelegten Reihenfolge

---

4 In den gängigen Vorgehensweisen werden entweder ab dem Startpunkt alle Adressen bis zur geforderten Anzahl aufgelistet oder nur jede x-te (z.B. jede dritte oder fünfte) Adresse bis zur geforderten Anzahl.

auflisten. Lebt nur eine Person der Grundgesamtheit im Haushalt, ist kein weiterer Auswahlschritt notwendig. In allen anderen Fällen muß mit Hilfe eines Zufallsverfahrens die zu befragende Person ermittelt werden. In den ALLBUS-Erhebungen wurde den Interviewern dafür jeweils eine ‚Kish Tabelle' (resp. ‚Schwedenschlüssel') vorgegeben, bei der die Befragungsperson anhand einer vorbereiteten Zufallsziffernreihe ausgewählt wird.[5]

Das ADM-Design wurde in neun ALLBUS-Erhebungen verwendet: 1980, 1982, 1990 und 1998 als Adress-Random, 1984, 1986 und 1988 als Random-Route und 1991 und 1992 als Standard-Random. In den verbleibenden drei Erhebungen 1994, 1996 und 2000 dagegen wurden Stichproben aus Einwohnermelderegistern gezogen, ein Vorgehen, das aufgrund seines zeitlichen und finanziellen Mehraufwandes bei bevölkerungsrepräsentativen face-to-face Umfragen nur selten zum Einsatz kommt (vgl. Koch 1997).

Bei diesem Verfahren wird in einer zweistufigen geschichteten Zufallsauswahl zunächst eine Stichprobe von Gemeinden gebildet und anschließend werden in den Gemeinden Personenadressen aus den Einwohnermelderegistern zufällig ausgewählt. Die Bruttostichprobe wird damit völlig unabhängig von der Feldarbeit erstellt, den Interviewern werden die zu befragenden Personen mit Namen und Adresse vorgegeben. Die Hauptvorteile im Vergleich zum ADM-Design sind die Minimierung des Interviewereinflusses bei der Auswahl der Befragungspersonen, die Verfügbarkeit von Registerinformationen (Alter, Geschlecht, deutsche Staatsangehörigkeit: ja/nein) für Teilnehmer und Nichtteilnehmer der Umfrage sowie designbedingt gleiche Auswahlwahrscheinlichkeiten für die Zielpersonen.

3. Realisierte Interviews pro Interviewer

Beginnen wir nun mit Informationen zu den zentralen Akteuren der Feldarbeit, den Interviewern. Sie stellen die wichtigste – und oft einzige[6] – Verbindung zwischen dem Forscher und den Befragten dar (vgl. Groves/Couper 1998: 191ff.). Im ALLBUS waren in den einzelnen Erhebungen zwischen 285 (ALLBUS 2000) und 719 (ALLBUS 1992) Interviewer an der Realisierung der jeweils 3.000 bis 3.500 Interviews beteiligt (vgl.

---

5 Werden alle Auswahlvorschriften dieses Verfahrens korrekt eingehalten, erhält man eine Stichprobe, bei der jeder Haushalt die gleiche Chance hat, in die Befragung einbezogen zu werden. Die Auswahlchance des einzelnen Haushaltsmitgliedes hängt von der Haushaltsgröße ab. Dies hat zur Folge, dass bei der Auswertung der Daten auf Personenebene vom Design her eine Gewichtung mit der – um die nicht zur Grundgesamtheit zählenden Haushaltsmitglieder reduzierten – Haushaltsgröße notwendig ist, die sogenannte ‚Transformation' (zur Problematik dieser Gewichtung vgl. Hartmannn/Schimpl-Neimanns 1992: 319ff.).

6 Zumeist kommt als zusätzlicher Kommunikationsweg lediglich ein Anschreiben des Forschers an alle Zielpersonen in Betracht, vgl. z.B. Lynn et al. 1998.

Tabelle 2).[7] Damit lag die Zahl der von jedem Interviewer im Durchschnitt durchgeführten Interviews zwischen 4,9 (1991, 1992) und 11,0 (2000).

Warum die Anzahl der Interviewer und deren durchschnittlicher „workload" zwischen den Studien variiert, kann hier nicht im einzelnen geklärt werden. Die Entscheidung über die Zahl der einzusetzenden Interviewer wird letztendlich vom Umfrageinstitut getroffen. Die Größe des Interviewerstabes, dessen Auslastung, aber auch die spezifischen Anforderungen der konkreten Studie spielen bei Entscheidungen über den Interviewereinsatz eine Rolle. Zu bedenken ist, dass in einem bestimmten Zeitraum von einem Institut in der Regel mehrere Studien parallel bearbeitet werden. Bei größeren Organisationen können durchaus ständig 15 bis 30 Projekte zeitgleich im Feld sein (vgl. Blyth 1998: 574), so dass die Planung und Steuerung des Interviewereinsatzes eine komplexe Aufgabe darstellt.

Tabelle 2: Realisierte Interviews pro Interviewer*

Jahr	Anzahl realisierter Interviews	Anzahl Interviewer	Realisierte Interviews pro Interviewer (Mittelwert)	Maximale Zahl der Interviews pro Interviewer
1980	2955	430	6,9	24
1982	2991	415	7,2	48
1984	3004	438	6,9	30
1986	3095	611	5,1	31
1988	3052	324	9,4	145
1990	3051	364	8,4	58
1991	3058	625	4,9	20
1992	3548	719	4,9	22
1994	3450	608	5,7	23
1996	3518	451	7,8	52
1998	3234	369	8,8	55
2000	3138	285	11,0	78

* nur Interviewer mit mindestens einem realisierten Interview
Datenbasis: ALLBUS 1980-2000

Für die Zukunft ist davon auszugehen, dass mit der zunehmenden Verbreitung von CAPI die Interviewerstäbe zahlenmäßig eher kleiner werden, die verbleibenden Interviewer

7 Beteiligt heißt in diesem Fall: die betreffenden Interviewer haben mindestens ein Interview realisiert. Daneben haben in jeder Erhebung noch eine Reihe weiterer Interviewer mitgearbeitet, ohne jedoch wenigstens ein Interview zu erzielen. Deren Anteil betrug in den einzelnen Erhebungen zwischen einigen wenigen Interviewern und circa 15 Prozent aller Interviewer.

jedoch in größerem zeitlichen Umfang der Interviewertätigkeit nachgehen als bisher. Auf diese Weise können die Umfrageinstitute die hohen Investitionskosten (für Laptops, Software, etc.) bei CAPI am ehesten amortisieren (vgl. Collins et al. 1998: 42)[8]. Unter methodischem Blickwinkel ist eine solche Entwicklung allerdings nicht unproblematisch. Wenn Interviewer homogenisierende Effekte auf die resultierenden Daten haben, dann können in ähnlicher Weise Designeffekte (mit der Folge einer Reduzierung der effektiven Stichprobengröße) auftreten wie aufgrund der räumlichen Klumpung einer Stichprobe (vgl. Schnell 1997: 272ff.). Diese Designeffekte werden umso größer ausfallen, je höher die Zahl der realisierten Interviews pro Interviewer ist (vgl. Fowler/Mangione 1990: 25ff.; Groves 1989: 360ff.).

Im übrigen ist darauf hinzuweisen, dass in manchen ALLBUS-Studien Interviewer vereinzelt weit mehr als die durchschnittliche Zahl von Interviews realisiert haben (vgl. Tabelle 2). Zwar wurde bei den Erhebungen des ALLBUS normalerweise eine Obergrenze für die Zahl der Interviews pro Interviewer definiert (diese betrug nie mehr als 50 Interviews pro Interviewer). Dies verhinderte jedoch nicht, dass in Einzelfällen dieses Limit überschritten wurde. In sechs der zwölf Erhebungen betrug die maximale Zahl von Interviews pro Interviewer zwischen 20 und 31, in vier Umfragen lag das Maximum bei 48 bis 58 Interviews. Höhere Werte weisen der ALLBUS 2000 (Maximum 78 Interviews) und insbesondere der ALLBUS 1988 auf. In dieser Erhebung hat ein Interviewer insgesamt 145 Interviews und damit knapp 5 Prozent aller Interviews der Studie durchgeführt. Auch wenn man in Rechnung stellt, dass bei der Feldarbeit mitunter Kompromisse zwischen dem theoretisch wünschenswerten und dem praktisch einlösbaren Vorgehen gemacht werden müssen, erscheint ein solch hoher Wert vor dem Hintergrund der oben angesprochenen Problematik von Interviewereffekten nur schwer tolerabel.

4. Demographische Struktur der Interviewer

Was wissen wir über die im ALLBUS eingesetzten Interviewer? Für die Erhebungen 1984 bis 1990 und 1994 bis 2000 liegen Informationen über das Geschlecht und das Alter, für die Umfragen 1994 bis 2000 zusätzlich Informationen zum Bildungsabschluß vor. Tabelle 3 gibt die Verteilungen der drei Interviewermerkmale wieder. Neben den Ergebnissen auf Ebene der Interviewer sind die Resultate auf Ebene der Interviews aufgeführt. Bei letzteren geht jeder Interviewer mit einem Gewicht proportional zur Zahl seiner durchgeführten

8 Collins et al. nehmen im übrigen an, dass sich mit der weiteren Verbreitung von CAPI auch die Beziehung zwischen Interviewer und Umfrageinstitut, incl. deren vertraglicher Gestaltung, ändern wird: die Interviewertätigkeit bietet dann unter Umständen die Perspektive einer richtigen Berufslaufbahn (Stichwort: „interviewing as a career“).

Interviews in die Analyse ein. Außerdem sind zum Vergleich die Verteilungen für die Befragten der betreffenden Erhebungen dokumentiert.

Wie Tabelle 3 belegt, unterscheiden sich die Ergebnisse auf Ebene der Interviewer und auf Ebene der Interviews nur unwesentlich. Dies bedeutet, dass die betreffenden Kategorien von Interviewern im Durchschnitt ungefähr gleich viele Interviews realisiert haben.[9] Im folgenden werden wir uns im Text jeweils auf die Ergebnisse für die Ebene der Interviews beziehen.

Tabelle 3: Demographische Merkmale von Interviewern und Befragten

	Interviewer* (Ebene der Interviewer)	**Interviewer* (Ebene der Interviews)**	**Befragte**
Geschlecht:**	**%**	**%**	**%**
Männlich	66,4	67,9	47,7
Weiblich	33,6	32,1	52,3
N	3435	25321	25542
Alter:**	**%**	**%**	**%**
18-29 Jahre	14,8	12,3	20,7
30-44 Jahre	32,2	29,3	28,9
45-59 Jahre	34,0	37,0	24,4
60 Jahre u. älter	19,0	21,4	26,0
N	3431	25418	25500
Durchschnittsalter	45,8 Jahre	47,2 Jahre	46,3 Jahre
Bildung:***	**%**	**%**	**%**
Bis Hauptschulabschluß	16,8	17,5	47,7
Mittlere Reife/ Realschulabschluß	33,3	35,0	29,7
Fachhochschulreife/ Abitur	17,2	15,7	9,6
Fachhochschul-/ Hochschulabschluß	32,7	31,8	13,1
N	1695	13187	13168

* Nur Interviewer mit mindestens einem realisierten Interview
** Datenbasis ALLBUS 1984-1990, 1994-2000
*** Datenbasis ALLBUS 1994-2000

9 Die Vermutung liegt nahe, dass die betreffenden Kategorien von Interviewern sich damit grosso modo auch kaum in ihren Erfolgs-, sprich Ausschöpfungsquoten unterscheiden. Zutreffend ist diese Folgerung dann, wenn die Interviews der unterschiedlichen Kategorien von Interviewern aus einem vergleichbar großen Bruttoansatz resultieren. Da entsprechende Informationen mit einer Ausnahme (ALLBUS 1986, vgl. Anmerkung 14) nicht vorliegen, kann dies hier nicht geprüft werden.

Zwei Drittel der ALLBUS-Interviews wurden von männlichen und ein Drittel von weiblichen Interviewern durchgeführt. Das Durchschnittsalter der Interviewer und der Befragten unterscheidet sich mit 47 bzw. 46 Jahren kaum, allerdings ist die jüngste und älteste Altersgruppe unterproportional, die Gruppe der 45-59jährigen dagegen überproportional unter den Interviewern vertreten. Die Interviewer verfügen über ein höheres Bildungsniveau als die Befragten: nahezu 50 Prozent aller Interviews wurden von Interviewern mit Fachhochschulreife/Abitur bzw. einem Fachhochschul- oder Hochschulabschluß durchgeführt, während von den Befragten nur 23 Prozent über einen solchen Bildungsabschluß verfügen.[10]

Dieses Profil der ALLBUS-Interviewer entspricht in der Grundtendenz den verfügbaren Informationen über die Interviewerstäbe bei anderen Erhebungen (vgl. zum Sozio-oekonomischen Panel z.B. Schräpler/Wagner 2001). Im internationalen Vergleich ist allerdings die überproportionale Beteiligung männlicher Interviewer auffällig. Nach Groves/Couper (1998: 196) dominieren bei face-to-face Befragungen weibliche Interviewer, so z.B. in den USA, aber auch in Kanada oder den Niederlanden. Dem ALLBUS vergleichbare Befunde werden dagegen für Alter und Bildung mitgeteilt. Für mehrere Studien aus den USA berichten die Autoren ein mittleres Alter der Interviewer von circa 50 Jahren sowie ein überdurchschnittliches Bildungsniveau.

Eine Differenzierung der Ergebnisse nach Erhebungszeitpunkt sowie nach West- und Ostdeutschland (nicht tabellarisch ausgewiesen) ergibt bei der Geschlechterverteilung kaum Unterschiede zwischen den Jahren.[11] Festzuhalten ist jedoch, dass in Ostdeutschland der ‚Männerüberschuß' bei den Interviewern stärker ausgeprägt ist als in Westdeutschland. Das Durchschnittsalter der Interviewer nimmt über die Zeit zu: Waren die ALLBUS-Interviewer in den 80er Jahren im Mittel zwischen 40 und 45 Jahren alt, sind sie in den 90ern circa 50 bis 55 Jahre alt.[12] In Ostdeutschland sind die Interviewer durchschnittlich zwei bis sieben Jahre älter als in Westdeutschland. Weit gravierender ist jedoch der Unterschied zwischen West und Ost im Bildungsniveau der Interviewer: In Ostdeutschland stammen circa 60 Prozent aller Interviews von Interviewern mit einem Fachhochschul- oder Hochschul-

10 Inwieweit die demographische Zusammensetzung des Interviewerstabes von den Umfrageinstituten bewußt gesteuert wird, ist nicht bekannt. Das größte Umfrageinstitut in Deutschland – die GfK in Nürnberg – gibt als Zielsetzung an, den Interviewerstab entsprechend dem soziodemographischen Aufbau der Bevölkerung zu strukturieren (vgl. Niehoff 1998: 54).

11 Auch bei dieser differenzierten Betrachtung unterscheiden sich die Resultate für die beiden Ebenen der Interviewer und der Interviews nicht gravierend. Im Text werden wieder die Ergebnisse auf Ebene der Interviews berichtet.

12 Z.T. mag dies daher rühren, dass über die Zeit die gleichen Interviewer von den Umfrageinstituten für die Durchführung der ALLBUS-Studien eingesetzt wurden.

abschluß, im Westen sind es weniger als 20 Prozent. In dem beobachtbaren Zeitraum zwischen 1994 und 2000 haben sich diese Werte kaum verändert.[13]

Warum ist die Zusammensetzung des Interviewerstabes überhaupt von Interesse? Zum einen kann das Teilnahmeverhalten der Zielpersonen beeinflußt werden. Verschiedene Kategorien von Interviewern können unterschiedlich hohe Ausschöpfungsquoten erzielen[14] oder aber bei unterschiedlichen Befragtengruppen besonders erfolgreich sein. In Deutschland wurde die Bedeutung der demographischen Komposition der Interviewerstäbe u.a. im Kontext der Media-Analyse diskutiert (vgl. Context 1993). Ausgangspunkt war die Beobachtung, dass die an dieser Gemeinschaftsuntersuchung beteiligten Umfrageinstitute unterschiedliche Reichweiten für die erfragten Zeitschriften ermittelten. Man vermutete, dass die Abweichungen mit der unterschiedlichen Strukturierung der Interviewerstäbe der Institute zusammenhängen. Bei Analysen des Teilnahmeverhaltens stellte man ein „Homogamie-Prinzip" der Gesellung fest: Männer haben eher Männer interviewt, Frauen eher Frauen, Ältere eher Ältere, usw. Erklärbar ist ein solches Ergebnis dadurch, dass bei einer Begegnung unter Fremden – wie es die Kontaktaufnahme für ein Interview typischerweise ist – eine Person mit ähnlichen Merkmalen sympathischer eingeschätzt wird und diese Sympathie die Kooperationsneigung erhöht (vgl. Groves/Couper 1998: 34). Auch im ALLBUS existieren Indizien für einen entsprechenden Zusammenhang zwischen Interviewer- und Befragtenmerkmalen.

Korreliert man in den einzelnen Erhebungen – für West- und Ostdeutschland getrennt – jeweils Geschlecht, Alter und Bildung der Interviewer mit dem betreffenden Merkmal der Befragten, dann erweisen sich 13 der 32 betrachteten Zusammenhänge als signifikant. In allen 13 Fällen besteht eine schwache positive Korrelation, das heißt auch hier läßt sich das in der Media-Analyse festgestellte „Homogamie-Prinzip" beobachten.[15] Auffällig

---

13 Die Ursache für das hohe Bildungsniveau der ostdeutschen Interviewer ist dabei nicht ganz klar. Von einem Umfrageinstitut wurde auf eine entsprechende Anfrage die Vermutung geäußert, dass in Ostdeutschland viele Personen, die aufgrund des Transformationsprozesses arbeitslos wurden (insbesondere auch Lehrer), sich als Interviewer betätigen.

14 Für den ALLBUS 1986 berichtet z.B. Koch (1991), dass ältere Interviewer eine höhere Ausschöpfung erzielen als jüngere und dass weibliche Interviewer erfolgreicher als männliche sind. Die Wirkung solcher fixen Interviewermerkmale erfolgt z.T. vermutlich vermittelt über Faktoren wie die Interviewererfahrung, die Verfügung über ein breites Repertoire von Handlungsstrategien, u.ä., so dass bei Einführung entsprechender Kontrollvariablen die Zusammenhänge reduziert werden (vgl. Groves/Couper 1998: 198).

15 Es bleibt weiterführenden Analysen vorbehalten zu klären, inwieweit diese Ergebnisse auch bei Beachtung möglicher methodischer Einwände Bestand haben. Zwei Kritikpunkte können gegen das gewählte Vorgehen geltend gemacht werden: Zum einen läßt die hier vorgenommene Analyse die Clusterung der Beobachtungen innerhalb der ‚workloads' der Interviewer außer Betracht, so dass die

dabei ist, dass die Zusammenhänge häufiger bei den ADM-Studien als bei den Registerstichproben auftreten: Bei ersteren erweisen sich zehn von vierzehn Zusammenhängen als signifikant, bei letzteren drei von achtzehn. Eventuell spielen für dieses Ergebnis die größeren Freiräume der Interviewer bei der Haushalts- und Zielpersonenauswahl im ADM-Design eine Rolle. Diese Hypothese deckt sich mit den Ergebnissen von Sodeur (1997: 81), der für mehrere ALLBUS-Stichproben anhand interner Prüfkriterien analysierte, inwiefern Abweichungen von der idealen Realisierung der Wahrscheinlichkeitsauswahl zu beobachten sind. Dabei schnitt der auf Basis einer Registerstichprobe durchgeführte ALLBUS 1994 besser ab als die ADM-basierten Studien, was Sodeur auf die geringeren Spielräume der Interviewer bei der Befragtenauswahl zurückführt.

Tabelle 4: Korrelation zwischen Interviewer- und Befragtenmerkmalen

Jahr	Erhebungsinstitut	Stichprobenverfahren	West			Ost		
			Geschlecht	Alter	Bildung	Geschlecht	Alter	Bildung
			Pearson's r			Pearson's r		
1984	GETAS	ADM	.09***	.15***				
1986	Infratest	ADM	.05**	.05**				
1988	GETAS	ADM	.05**	.09***				
1990	Infas	ADM	.02	.12***				
1994	Infratest	Register	.03	.02	.09***	.03	-.02	.05
1996	Infratest	Register	.01	.02	.04	-.01	.02	.06*
1998	GETAS	ADM	.02	.12***	.14***	.07*	.06	.05
2000	Infratest	Register	.03	.00	.05*	-.01	-.02	-.01

*= $p \leq .05$, ** = $p \leq .01$, *** = $p \leq .001$
Datenbasis: ALLBUS 1984-1990, 1994-2000

Neben dem Teilnahmeverhalten kann auch das Antwortverhalten der Befragten durch demographische Charakteristika der Interviewer beeinflußt werden (vgl. Reinecke 1998,

Berechnung der Standardfehler möglicherweise verzerrt ist. Zum anderen erfolgte die Zuweisung der Interviewer auf die Befragten nicht zufällig. Es kann damit nicht völlig ausgeschlossen werden, dass ältere Interviewer häufiger in Gegenden mit einem höheren Altersdurchschnitt tätig waren; Interviewer mit höherem Bildungsabschluß eher in „gehobenen" Wohnvierteln mit überdurchschnittlichem Bildungsniveau, usw. Sehr wahrscheinlich erscheint dies allerdings nicht. Faktisch ist eine Zufallszuweisung der Interviewer auf die Befragten in bundesweiten face-to-face Umfragen zwar schwieriger als bei Telefonumfragen, aber nicht prinzipiell unmöglich. Ein Vorgehen, um Interviewereffekte und Effekte der Region trennen zu können, besteht etwa darin, dass Paare von Interviewern sich die Interviews in jeweils zwei Befragungsclustern teilen (vgl. Groves 1989: 361f und Schnell/Kreuter 2000).

der einen Überblick über einschlägige neuere Studien gibt).[16] Potentiell möglich ist dies insbesondere dann, wenn zwischen einem Interviewermerkmal und dem Inhalt einer Frage ein Zusammenhang besteht. Das bekannteste Beispiel in dieser Hinsicht sind Fragen, die sich mit Rassenbeziehungen beschäftigen. Für die USA wurde wiederholt festgestellt, dass sich die Antwortverteilungen zwischen schwarzen und weißen Interviewern bei Fragen, die sich mit der Diskriminierung der schwarzen Bevölkerung beschäftigen, unterscheiden (vgl. Fowler/Mangione 1990: 98ff.; Groves 1989: 398ff.).[17] Schanz/Schmidt (1984) allerdings fanden bei ihrer Analyse ausgewählter Fragen des ALLBUS 1980 nur inkonsistente und schwache Effekte soziodemographischer Merkmale der Interviewer auf das Antwortverhalten der Befragten.[18]

5. Zeitpunkt der Interviewdurchführung

Vorgaben für das Kontaktverhalten der Interviewer stellen einen wichtigen Bestandteil des Designs einer Umfrage dar; sie spielen insbesondere für das Erreichen der Zielpersonen eine entscheidende Rolle. Es ist gut belegt, dass Umfragen, die häufige Kontaktversuche der Interviewer, an unterschiedlichen Wochentagen und zu unterschiedlichen Tageszeiten, verteilt über einen längeren Zeitraum, vorsehen, überdurchschnittliche Ausschöpfungsergebnisse aufweisen (vgl. Groves/Couper 1998: 79ff.). Leider verfügen wir im ALLBUS nur für wenige Erhebungen über detaillierte Daten zur Zahl und zeitlichen Terminierung der Kontaktversuche der Interviewer. In allen Studien (mit Ausnahme von 1991) liegen jedoch bei erfolgreicher Interviewdurchführung Angaben zum Wochentag vor. Wie man Tabelle 5 entnehmen kann, wurden 84 Prozent aller ALLBUS-Interviews unter der Woche zwischen Montag und Freitag realisiert. Zwischen den einzelnen Werktagen bestehen dabei nur geringe Unterschiede: Die meisten Interviews fanden an einem Montag, Dienstag oder Mittwoch statt. Von Donnerstag bis Samstag nimmt der Anteil kontinuierlich etwas ab. Der Sonntag scheint für die Interviewertätigkeit weitgehend tabu zu sein, nur 3,6 Prozent aller ALLBUS-Interviews wurden an einem Sonntag realisiert.

---

16 Es ist zu vermuten, dass der zugrundeliegende Mechanismus für solche Effekte eine Form sozial erwünschten Antwortverhaltens (z.B. im Sinne einer Angleichung an die vermutete Einstellung des Interviewers) darstellt.

17 Z.T. macht man sich diesen Mechanismus auch bewußt zunutze, indem ein „race-“ oder auch „gender-matching“ zwischen Interviewern und Befragten unter der Annahme vorgenommen wird, dass bei Merkmalsgleichheit der Beteiligten validere Antworten erzielt werden (vgl. Groves 1989: 400ff.).

18 Schräpler/Wagner (2001) berichten ein Analysebeispiel für das Sozio-oekonomische Panel, bei dem Alter und Bildung der Interviewer einen signifikanten Effekt auf das subjektive Wohlbefinden der Befragten aufweisen.

Tabelle 5: Wochentag, an dem die Interviews realisiert wurden

Wochentag	%
Montag	17,8
Dienstag	18,2
Mittwoch	17,8
Donnerstag	15,8
Freitag	14,4
Samstag	12,4
Sonntag	3,6
N	34952

Datenbasis: ALLBUS 1980-2000, ohne ALLBUS 1991

Dieses Muster ist recht stabil, zwischen den Erhebungen bzw. zwischen West- und Ostdeutschland sind nur geringe Unterschiede zu beobachten (nicht tabellarisch ausgewiesen). Die größten Abweichungen bestehen darin, dass 1980 besonders viele Interviews samstags (18,2 Prozent) durchgeführt wurden, während 2000 dieser Anteil mit 6,8 Prozent besonders niedrig ist. Im Gegenzug wurden im Jahr 2000 mit 8,7 Prozent überdurchschnittlich viele Interviews an einem Sonntag gemacht.

Für die Erhebungen ab 1991 liegen auch Informationen zu den Anfangszeiten der Interviews vor (vgl. Tabelle 6). Etwa ein Fünftel aller Interviews der ALLBUS-Erhebungen 1991 bis 2000 wurden in West- und Ostdeutschland vormittags vor 12 Uhr begonnen, zwei Fünftel nachmittags zwischen 12 und vor 17 Uhr und die restlichen zwei Fünftel abends um 17 Uhr oder später.

Tabelle 6: Uhrzeit, zu der die Interviews begonnen wurden

Jahr	West				Ost			
	N	vor 12 Uhr	12 bis vor 17 Uhr	17 Uhr oder später	N	vor 12 Uhr	12 bis vor 17 Uhr	17 Uhr oder später
		%	%	%		%	%	%
1991	1500	12,0	38,1	49,9	1537	13,3	32,7	54,0
1992	2391	11,5	34,4	54,1	1148	14,3	30,6	55,1
1994	2264	15,4	45,4	39,2	1050	21,6	43,7	34,7
1996	2375	16,2	42,4	41,5	1108	23,9	42,5	33,6
1998	2167	22,9	44,7	32,4	1016	25,8	46,8	27,5
2000	1945	22,7	47,7	29,6	1044	31,1	44,3	24,5
Gesamt	12642	16,8	42,1	41,1	6903	21,0	39,4	39,6

Datenbasis: ALLBUS 1991-2000

Über die Jahre hat der Anteil der Interviews, die vormittags oder nachmittags durchgeführt worden sind, zu-, der Anteil der abends realisierten Interviews abgenommen. Während im ALLBUS 1991 und 1992 in West- und Ostdeutschland jedes zweite Interview um 17 Uhr oder später begonnen wurde, sind es im ALLBUS 2000 lediglich circa 30 Prozent in West- und 25 Prozent in Ostdeutschland. Kombiniert man die Informationen zu Wochentag und Uhrzeit, so ergibt eine Differenzierung zwischen ‚Wochenende', ‚Werktags tagsüber' und ‚Werktags abends' im wesentlichen eine Verlagerung der Interviews von ‚Werktags abends' zu ‚Werktags tagsüber', während der Anteil der Wochenend-Interviews kaum variiert. 1992 wurden in Westdeutschland 35,6 Prozent der ALLBUS-Interviews tagsüber zwischen Montag und Freitag durchgeführt, im Jahr 2000 waren es 57,9 Prozent. In Ostdeutschland stieg der betreffende Anteil von 33,1 Prozent auf 64,1 Prozent.

Stellen diese Ergebnisse ein Indiz dafür dar, dass Interviewer in immer geringerem Umfang für Interviews zur Abendzeit zur Verfügung stehen? Wenn dem so wäre, wäre dies eine problematische Entwicklung, denn der Abend gilt gemeinhin – auch in der Sicht der Interviewer – als günstigster Zeitraum für die Interviewdurchführung.[19] Eine generelle Tendenz zur Verlagerung der verfügbaren Interviewzeit auf tagsüber läßt sich jedoch aus den vorliegenden Ergebnissen nicht zwingend ableiten, da die studienspezifischen Informationen für den ALLBUS keineswegs „repräsentativ" für die Verwendung des gesamten Zeitbudgets der Interviewer sein müssen. Allerdings liefert ein Vergleich von Befragungsergebnissen der am ALLBUS 1986 und ALLBUS 2000 beteiligten Interviewer zusätzliche Hinweise auf einen entsprechenden Trend. In diesen Befragungen wurde u.a. ermittelt, an welchen Wochentagen und zu welchen Tageszeiten die Interviewer normalerweise versuchen, Interviews für Bevölkerungsbefragungen durchzuführen. Während im Hinblick auf die Wochentage zwischen 1986 und 2000 kaum Unterschiede bestehen (lediglich der Freitag wird im Jahr 2000 etwas seltener als 1986 erwähnt), sind bei den genannten Uhrzeiten klare Veränderungen zu erkennen. Die vorgegebenen Zeitspannen vormittags und nachmittags (vor 12 Uhr, 12-15 Uhr, 15-17 Uhr) finden 2000 bei mehr Interviewern Anklang als 1986; die Zunahme beträgt zwischen elf und vierzehn Prozentpunkten. Im Gegenzug nimmt der Anteil der Interviewer, die angeben, zwischen 17 und 20 Uhr zu interviewen, um sechs bis sechzehn Prozentpunkte ab. Nicht ganz in dieses Bild paßt, dass der

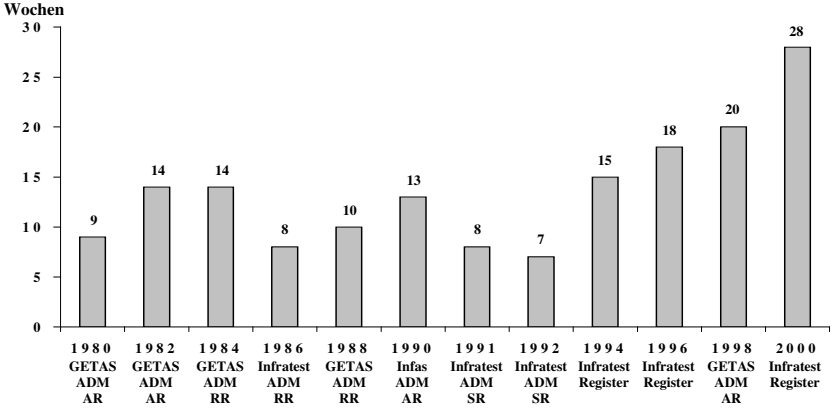
19 In von Infratest durchgeführten Interviewerbefragungen wurde u.a. erfaßt, welche Tageszeiten nach Meinung der Interviewer für Bevölkerungsbefragungen am günstigsten sind. 1984, 1989 und 1991 nannte jeweils die Mehrheit (zwischen 58 und 63 Prozent) der Infratest-Interviewer die Zeitspanne zwischen 17 und 19 Uhr (vgl. Bliesch o.J.). Schnell (1997: 226ff.) nimmt auf der Basis von Auswertungen von Zeitbudget-Daten an, dass in der Bundesrepublik die günstigste Zeit für ein Interview an Wochentagen zwischen 17 und 19.30 Uhr liegt. Groves/Couper (1998: 100ff.) zeigen für Studien aus den USA, dass Kontaktversuche zur Zielperson abends am erfolgreichsten sind.

Anteil der Interviewer, die berichten, auch nach 20 Uhr zu arbeiten, im Jahr 2000 um sechs Prozentpunkte höher als 1986 liegt.

6. Feldzeit

Betrachten wir nun die Feldzeit der ALLBUS-Umfragen. Rein theoretisch sollte die Länge der Feldzeit einer Umfrage so gewählt werden, dass sowohl schwieriger zu erreichende Befragungspersonen kontaktiert als auch weniger kooperationsbereite Zielpersonen z.B. mithilfe eines Wechsels des Interviewers doch noch zur Teilnahme motiviert werden können. In der Praxis spielen allerdings für die Länge der Feldzeit einer Umfrage eine Vielzahl von Faktoren eine Rolle – angefangen von designbedingten Unterschieden wie dem verwendeten Stichprobenverfahren bis zu konkreten studienspezifischen Einsatzbedingungen wie etwa der Auslastung des eingesetzten Interviewerstabes.

Abbildung 1: Feldzeit nach Erhebungsjahr: ALLBUS 1980-2000



Wie Abbildung 1 zeigt, erfolgte die Realisierung der jeweils 3.000 bis 3.500 Interviews des ALLBUS in sehr unterschiedlich langen Zeitspannen. Im Minimum wurden sieben Wochen (beim ALLBUS 1992), im Maximum 28 Wochen (beim ALLBUS 2000) Feldzeit benötigt. Dabei lassen sich einige charakteristische Unterschiede zwischen den verwendeten Stichprobenverfahren ausmachen. So wurde die Studie mit der kürzesten Feldzeit – der ALLBUS 1992 – nach dem ADM-Standard-Random-Verfahren durchgeführt. Dieses Verfahren ist speziell darauf angelegt, eine bestimmte Fallzahl in einer kurzen Zeit zu erzielen. Das Grundprinzip dabei ist, durch die Vorgabe einer Nettofallzahl zu realisierender Interviews und eines großzügig bemessenen bzw. zahlenmäßig nicht begrenzten Bruttos von bearbeitbaren Adressen Nachbearbeitungen weitgehend zu vermeiden. Die Feldarbeit wird sozusagen in einem Anlauf erledigt. Dabei besteht allerdings die Gefahr, dass leichter erreichbare und besonders kooperationsbereite Zielpersonen in der realisierten Stichprobe überproportional vertreten sind. Auf der anderen Seite wurde für die Umfrage mit der längsten Feldzeit – den ALLBUS 2000 – eine Stichprobe aus Einwohnermelderegistern verwendet. Hier werden den Interviewern die zu befragenden Personen mit Namen und Adresse vorgegeben. Durch diese feste Definition der Bruttostichprobe besteht keine Möglichkeit zur Substitution schwer erreichbarer oder wenig kooperationsbereiter Zielpersonen. Eine befriedigende Ausschöpfung und die geplanten (Netto-)Fallzahlen können bei einer solchen Stichprobe nur erzielt werden, wenn jede ausgewählte Adresse intensiv bearbeitet wird. Dies ist ohne zeitaufwendige Nachbearbeitungen kaum möglich.[20] Insgesamt lassen sich die ALLBUS-Erhebungen nach den eingesetzten Stichprobenverfahren in eine Rangfolge bringen, bei der die „weicheren", weniger geregelten und dafür mit mehr Ermessensspielraum für die Interviewer behafteten Verfahrensmodelle die kürzeste Feldzeit benötigen: Die beiden Umfragen nach dem Standard-Random-Modell weisen eine Feldzeit von durchschnittlich 7,5 Wochen auf, es folgen die drei Random-Route-Erhebungen (10,7 Wochen), die vier Adress-Random-Studien (14,0 Wochen) und die drei Registerstichproben (20,3 Wochen).

Neben solchen in der Verfahrenslogik begründeten Unterschieden spielen für die Dauer der Datenerhebung jedoch weitere, in den studienspezifischen Einsatzbedingungen liegende Faktoren eine Rolle: die zeitliche Plazierung der Feldzeit (Ferienzeit), die Größe und Auslastung des eingesetzten Interviewerstabes, aber auch solch idiosynkratische Ereignisse wie ein Poststreik können die Feldzeit beeinflussen. So ist etwa die extrem

20 Daneben kann ein weiterer Aspekt für die Unterschiede in der Feldzeit zwischen ADM- und Registerstichproben eine Rolle spielen: die Wegestrecken für die Interviewer sind aufgrund der stärkeren Streuung der Adressen bei der Registerstichprobe länger als beim ADM-Design. Dadurch wird die konzentrierte Bearbeitung einzelner räumlicher Einheiten schwieriger, die Zahl der notwendigen Anfahrten in den Point erhöht sich.

lange Feldphase des ALLBUS 2000 nicht ausschließlich auf die Verwendung einer Registerstichprobe zurückzuführen. Eine solche Stichprobe kam auch beim ALLBUS 1994 zum Einsatz, bei dem die Feldzeit mit 15 Wochen nur knapp halb so lang war. Ein wesentlicher Unterschied zwischen beiden Erhebungen bestand freilich darin, dass der ALLBUS 2000 erstmals computerunterstützt durchgeführt wurde. Dadurch war die Zahl der beteiligten Interviewer nur halb so groß wie 1994; überdies war der kleinere CAPI-Stab im Frühjahr 2000 aufgrund des hohen Studienaufkommens bei Infratest stark belastet. Zusätzlich spielte eine Rolle, dass in der 2000er Erhebung aufgrund des schlechten Ergebnisses nach der regulären Feldzeit eine vierwöchige Phase angeschlossen wurde, in der eine Reihe von besonders gering ausgeschöpften Sample Points (insgesamt circa zehn Prozent aller Points) ausgetauscht und neu bearbeitet wurden (vgl. Koch et al. 2001: 53ff.).

Einen etwas detaillierteren Einblick in den Feldverlauf bietet eine Betrachtung auf Ebene der primären Stichprobeneinheiten[21] – bei den ADM-Studien sind dies Wahlbezirke, bei den Registerstichproben jeweils Adressklumpen von zehn Adressen, die innerhalb der Gemeinden aufgrund ihrer räumlichen Nähe zusammengefaßt und den Interviewern zur Bearbeitung vorgegeben wurden. Die Zahl dieser Klumpen variierte im ALLBUS zwischen 600 (1998) und 752 (1992).[22] Im Durchschnitt wurden in den Erhebungen pro Klumpen zwischen 4,6 (1986) und 5,4 (1994, 1996, 1998) Interviews realisiert. Darauf hinzuweisen ist dabei allerdings, dass das Ausgangsbrutto in den Studien nicht gleich groß war: es betrug „offiziell“ zwischen sieben und zehn Adressen pro Klumpen bzw. es war im Fall der Standard-Random-Studien 1991 und 1992 durch eine Obergrenze von 12 bzw. 15 Adressen definiert. Normalerweise wird ein solcher Klumpen von einem Interviewer bearbeitet; ein zweiter Interviewer wird lediglich dann tätig, wenn im Zuge der Nachbearbeitung ein Interviewerwechsel vorgenommen wird. Der Anteil der Klumpen, die von mehr als einem Interviewer bearbeitet wurden, betrug in den Erhebungen zwischen 5,1 Prozent (1984) und 36,7 Prozent (1994).

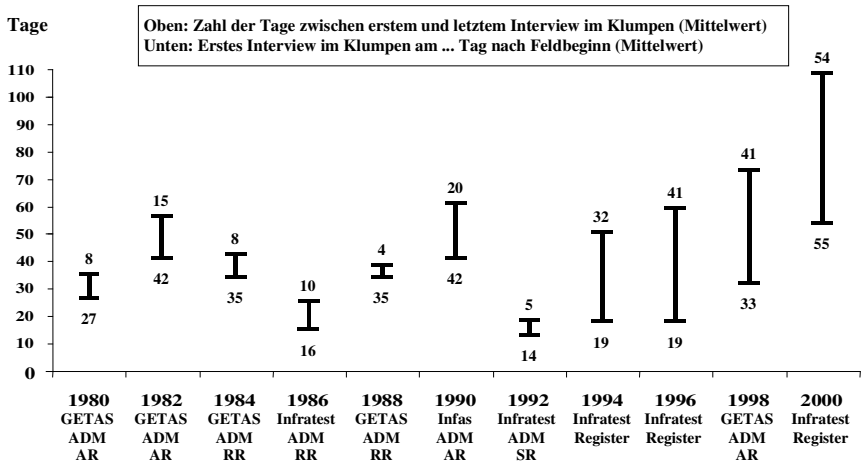
In Abbildung 2 sind zwei Informationen abgetragen: die durchschnittliche Zahl der Tage, bis das erste Interview im Klumpen vorlag, und die durchschnittliche Zahl der Tage zwischen erstem und letztem Interview im Klumpen. Richtet man die Aufmerksamkeit zunächst auf die Tage bis zum ersten Interview, so fällt auf, dass der offizielle Start der Feldarbeit keinesfalls gleichzusetzen ist mit dem Beginn der Bearbeitung sämtlicher

21 Vgl. zu dieser Herangehensweise auch Schnell 1997 (229ff.), der die Feldarbeit für den ALLBUS 1988 näher beleuchtet.

22 Die genannten Zahlen beziehen sich auf die Klumpen, in denen mindestens ein Interview realisiert wurde (auch Abbildung 2 liegt diese Abgrenzung zugrunde). Die Zahl der Klumpen ohne realisiertes Interview lag in den einzelnen Erhebungen zwischen zwei (1996) und 48 (1992).

Klumpen. In den einzelnen ALLBUS-Erhebungen dauerte es im Durchschnitt zwischen 14 (1992) und 55 (2000) Tagen, bis das erste Interview in den Klumpen realisiert wurde. Die Studien unterscheiden sich also deutlich darin, wie intensiv in die Feldarbeit bei Feldbeginn eingestiegen wurde.[23]

Abbildung 2: Feldbeginn und Feldzeit auf Klumpen-Ebene: ALLBUS 1980-2000*



* ohne ALLBUS 1991, bei dem keine Angaben zum Datum des Interviews verfügbar sind

Dabei sind Unterschiede zwischen den Umfrageinstituten erkennbar: Bei den von GETAS bzw. Infas durchgeführten Erhebungen wurde das erste Interview im Klumpen durchschnittlich nach 27 bis 42 Tagen realisiert; das heißt im Mittel lag erst vier bis sechs Wochen nach Feldbeginn ein greifbares Resultat der Feldarbeit vor. Schneller war dagegen – mit einer Ausnahme – Infratest: im Durchschnitt wurde das erste Interview in den Klumpen nach 14 bis 19 Tagen erbracht. Dabei machte es keinen Unterschied, ob eine ADM-Stichprobe (wie z.B. im Jahr 1986) oder eine Registerstichprobe (wie z.B. 1994) zugrundelag. Die Ausnahme bildet der von Infratest durchgeführte ALLBUS 2000: im Durchschnitt lag hier das erste Interview in den Klumpen nach acht Wochen vor. Dieser lange Zeitraum

23 Darauf hinzuweisen ist, dass die hier vorgenommene Analyse sich auf die Interviewdurchführung bezieht. In vielen Fällen ist es möglich, dass dieser bereits Kontaktversuche zu früheren Zeitpunkten vorausgingen.

hängt mit der schon erwähnten geringeren Zahl von CAPI-Interviewern und deren großer Auslastung in dem betreffenden Zeitraum zusammen (deswegen wurde beispielsweise ein Viertel aller Adressen erst vier bis zehn Wochen nach Feldbeginn an die Interviewer zur Bearbeitung gegeben). Außerdem spielt der bereits angesprochene Austausch und die Neubearbeitung von Befragungsklumpen am Ende der Feldzeit eine Rolle: ohne diesen reduziert sich der Mittelwert der Tage bis zum ersten Interview von 55 auf 42.

Die Zahl der Tage zwischen erstem und letztem Interview in den einzelnen Klumpen vermittelt einen Eindruck von der effektiven Bearbeitungszeit, die für die Klumpen aufgewendet wurde. Man erkennt in Abbildung 2 zunächst, dass die faktischen Bearbeitungszeiten der Klumpen weit kürzer sind als die Feldphase insgesamt. Sie liegen zwischen durchschnittlich vier (1988) und 54 (2000) Tagen und betragen damit zwischen 1/17 (1988) und 1/3 (1996) der jeweiligen Feldzeit. Dies bedeutet, dass die Länge der Feldzeit einer Erhebung zu einem erheblichen Teil davon abhängig ist, inwieweit die Befragungsklumpen zeitgleich oder aber zeitversetzt bearbeitet werden.

Bei einigen der – face-to-face durchgeführten – ALLBUS-Erhebungen ist die faktische Feldphase kaum länger als die bei Telefonumfragen übliche Zeitspanne von wenigen Tagen. Den (Negativ-) Rekord halten in dieser Hinsicht die ALLBUS-Erhebungen 1988 und 1992. Im Durchschnitt lagen hier gerade vier bzw. fünf Tage zwischen erstem und letztem Interview in den Klumpen. In 77 bzw. 74 Prozent aller Klumpen wurden alle Interviews innerhalb einer Woche, in 39 bzw. 8 Prozent aller Klumpen sogar an einem Tag durchgeführt. Man kann mit einiger Berechtigung bezweifeln, ob bei diesen Erhebungen die faktische Datenerhebungsphase damit lang genug war, um alle Zielpersonen wirklich erreichen zu können (zum Zusammenhang von Länge der Feldzeit und Erreichbarkeit von Zielpersonen vgl. Groves/Couper 1998: 272ff.).

Auch in der Frage der effektiven Bearbeitungszeit sind wieder Unterschiede zwischen den eingesetzten Stichprobenverfahren zu beobachten. Die längste Feldzeit auf Klumpenebene weisen die drei Registerstichproben mit einer durchschnittlichen Bearbeitungszeit von 42,3 Tagen auf. An zweiter Stelle folgen die ADM-Stichproben mit einem separaten Adressenvorlauf mit 21,0 Tagen und das Schlußlicht bilden die Random-Route- bzw. Standard-Random-Studien mit 7,3 bzw. 5,0 Tagen. Dass gerade zwischen diesen beiden am weitesten verbreiteten Verfahrensmodellen kaum Unterschiede in der effektiven Bearbeitungszeit zu verzeichnen sind, ist womöglich ein Hinweis darauf, dass die Interviewer in der Feldpraxis nur wenig zwischen beiden Vorgehensweisen unterscheiden: das Random-Route-Verfahren, das sich eigentlich durch die feste Vorgabe einer bestimmten Zahl zu bearbeitender Bruttoadressen auszeichnet, kann durch die undokumentierte Substitution von nicht erfolgreich bearbeiteten Adressen faktisch dem Standard-Random-Modell

(mit einem zahlenmäßig nicht begrenzten oder eher großzügig bemessenen Brutto) sehr stark angenähert sein. Auf die Problematik der Adressensubstitution wird am Schluß des Beitrags nochmals einzugehen sein.

7. Fallpreise

Eine Betrachtung der Feldarbeit ohne Berücksichtigung der entstehenden Kosten wäre unvollständig. Im folgenden wollen wir einen Blick auf die Fallpreise in den einzelnen ALLBUS-Erhebungen werfen. Diese wurden ermittelt, indem der bei dem Umfrageinstitut angefallene Nettopreis der Studie durch die Zahl der erzielten auswertbaren Interviews dividiert wurde. In die Fallpreise gehen damit neben den Feldkosten im engeren Sinne (wie Honorare und Spesen der Interviewer, Kosten von Feldsteuerung, Interviewerschulung und -kontrolle) auch alle weiteren Aufwendungen im Umfrageinstitut ein: für die Stichprobenbildung, Datenaufnahme und -prüfung, Projektleitung, usw.[24] In der Regel bilden die Feldkosten allerdings den größten Posten in der Kalkulation einer Studie.

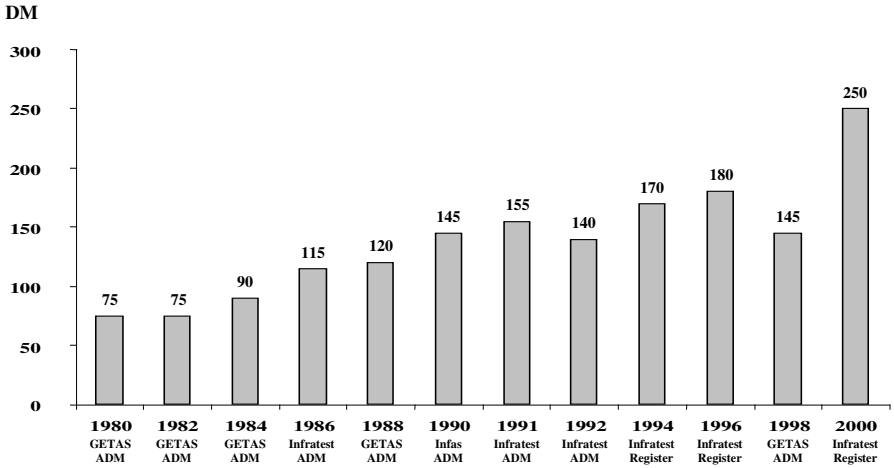
Abbildung 3 zeigt die Entwicklung der Fallpreise im ALLBUS zwischen 1980 und 2000. Während 1980 das ALLBUS-Interview durchschnittlich DM 75,- kostete, betrug der Preis zwanzig Jahre später DM 250,-. Dies würde eine Erhöhung der Fallpreise um den Faktor 3,3 bedeuten – eine weit größere Steigerung als sie etwa für die privaten Lebenshaltungskosten im gleichen Zeitraum zu verzeichnen ist (Faktor 1,6; berechnet nach Informationen in: Statistisches Bundesamt 2000: 328ff.). Zu berücksichtigen ist jedoch, dass die Erhebungen 1980 und 2000 sich in mehrerlei Hinsicht unterscheiden. So wurde der ALLBUS 1980 von GETAS durchgeführt, der ALLBUS 2000 von Infratest; 1980 wurde das ADM-Design verwendet, 2000 eine Registerstichprobe; 1980 existierte das ISSP noch nicht, 2000 wurde der ISSP-Fragebogen als Selbstausfüller im Anschluß an das ALLBUS-Interview erhoben.[25] Zieht man deshalb die in Hinblick auf das durchführende Umfrageinstitut, das verwendete Stichprobenverfahren sowie die Befragungsdauer vergleichbaren Erhebungen 1980 und 1998 heran – beide wurden von GETAS nach dem ADM-Design (mit Adressen-Vorlauf) bei einer durchschnittlichen Interviewdauer von jeweils 50 Minuten realisiert –, dann ist mit der Zunahme des Fallpreises von DM 75,- auf DM 145,- nurmehr eine knappe Verdoppelung zu beobachten.

---

24 Außen vor bleiben alle bei den Datenproduzenten ZUMA und ZA entstandenen Aufwendungen, etwa für die Fragebogenentwicklung, die weitergehende Datenprüfung und -dokumentation, usw.

25 Das International Social Survey Programme (ISSP) ist ein weltweiter Forschungsverbund, der seit 1985 jedes Jahr sozialwissenschaftliche Umfragen mit wechselnden Themenschwerpunkten durchführt. In Deutschland wird die ISSP-Umfrage zumeist als schriftlicher Selbstausfüller im Anschluß an das mündliche ALLBUS-Interview erhoben.

Abbildung 3: Fallpreise in den ALLBUS-Erhebungen 1980 – 2000 (o. MWSt.)



Generell ist davon auszugehen, dass Registerstichproben teurer als ADM-Stichproben sind, aber auch zwischen den verschiedenen Varianten des ADM-Designs existieren Preisunterschiede: am teuersten ist das Adress-Random mit der separaten Begehungsphase zur Ermittlung der Haushaltsadressen, darauf folgt das traditionelle Random-Route mit der festen Bruttovorgabe von zu bearbeitenden Adressen und am günstigsten ist das Standard-Random mit der Nettovorgabe zu realisierender Interviews. Neben solchen verfahrensbedingten Preisunterschieden sollte man überdies gewärtigen, dass auch zwischen verschiedenen Umfrageinstituten trotz vergleichbarer Leistungsangebote in der Preisgestaltung z.T. merkliche Differenzen zu verzeichnen sind.

Bedauerlicherweise existieren keine veröffentlichten Informationen zu den Fallpreisen anderer Bevölkerungsumfragen in der Bundesrepublik. Dennoch ist es wohl kaum verkehrt anzunehmen, dass z.B. der Fallpreis des ALLBUS 2000 mit DM 250,- am oberen

Limit dessen liegt, was in Deutschland bei ähnlichen Erhebungen bezahlt wird.[26] Es stellt sich die Frage, inwieweit dieser hohe Preis gerechtfertigt ist.

8. Qualität, Zeitbedarf und Preis

Selbstverständlich ist der Zusammenhang von Qualität und Preis in der Umfrageforschung zu komplex, als dass er im hier gegebenen Rahmen auch nur ansatzweise befriedigend behandelt werden könnte. Wenn das Thema zum Schluß dieses Beitrages dennoch aufgegriffen wird, dann um auf ein wichtiges praktisches Problem aufmerksam zu machen, dem sich eine qualitätsorientierte Surveyforschung in Deutschland gegenübersieht. Die Rede ist von der Zuverlässigkeit der Ermittlung der Ausschöpfungsquote, eines zentralen Qualitätsindikators in der Umfrageforschung.[27]

Das (Pseudo-)Motto eines amerikanischen Umfrageinstitutes faßt den Zusammenhang zwischen Qualität, Zeitbedarf und Preis so zusammen: „Quality, speed, price: pick two." (zit. nach Smith 1995: 167). Dieses Motto impliziert, dass bei hohen Preisen mit hoher Qualität in kurzer Zeit gerechnet werden kann. Im Gegenzug legt es nahe, dass bei niedrigen Preisen Abstriche in der Qualität oder der Geschwindigkeit in Kauf genommen werden müssen. In der Bundesrepublik stellt sich jedoch – zumindest bei oberflächlicher Betrachtungsweise – die Entscheidungssituation bei Bevölkerungumfragen anders dar: Zieht man als Qualitätsindikator die Ausschöpfungsquote heran, besteht (vermeintlich) die Wahl zwischen einer hohen Ausschöpfung, kurzen Feldzeit und einem niedrigen Preis einerseits oder einer niedrigen Ausschöpfung, langen Feldzeit und einem hohen Preis andererseits. Ersteres erhält man, wenn eine Stichprobe nach dem ADM-Random-Route-

---

26 Die ursprüngliche Kalkulation des ALLBUS 2000 sah einen etwas niedrigeren Fallpreis von DM 225,- vor. Dieser erhöhte sich dann allerdings auf DM 250,-, weil die angestrebte Nettofallzahl von 3.500 auswertbaren Interviews nicht realisiert wurde.
International nimmt der ALLBUS mit diesem Preis keine Spitzenposition ein. Nach einer mündlichen Auskunft von Tom Smith, dem Verantwortlichen für den amerikanischen General Social Survey (GSS), lag beispielsweise der Preis für ein Interview in der GSS-Erhebung 1996 um ein Mehrfaches höher.

27 „For many years, it seemed as if the design of the sample and the response rate were the only two indicators of a survey's quality." (Fowler/Mangione: 1990: 142). In der Praxis wird die Ausschöpfungsquote allerdings mitunter unreflektiert verwendet. Oft wird außer acht gelassen, dass sie lediglich ein Maß für die *potenzielle* Verzerrung von Daten durch einen Nonresponsebias ist. Die *tatsächliche* Stärke einer Verzerrung ist abhängig von der Höhe der Nonresponsequote einerseits und dem Ausmaß des Unterschiedes zwischen Teilnehmern und Nichtteilnehmern einer Umfrage andererseits, welches jeweils variablenspezifisch zu bestimmen ist (vgl. Groves/Couper 1998: 1ff.).

(bzw. Standard-Random-)Modell gewählt wird, letzteres stellt sich bei Verwendung einer Registerstichprobe ein.[28]

Zur Veranschaulichung dieser These sollen exemplarisch vier Bevölkerungsumfragen betrachtet werden, die von ZUMA in der jüngeren Vergangenheit betreut wurden und die alle für sich in Anspruch nehmen, hohen Qualitätsstandards zu genügen. Dies sind die beiden ALLBUS-Erhebungen 1996 und 2000, die auf Basis einer Registerstichprobe durchgeführt wurden, und der Wohlfahrtssurvey 1998 sowie der Sozialwissenschaften-Bus 1997, bei denen eine ADM-Stichprobe nach dem Random-Route-Modell zum Einsatz kam. Wie Tabelle 7 zeigt, beträgt bei dem auf Basis einer Registerstichprobe durchgeführten ALLBUS 2000 die Ausschöpfung lediglich 49 Prozent – und das bei einer Feldzeit von 28 Wochen und einem Fallpreis von DM 250,-. Etwas besser fällt das Ergebnis beim ALLBUS 1996 aus: hier wurden in 18 Wochen Feldzeit 54 Prozent Ausschöpfung erzielt; das einzelne Interview kostete durchschnittlich DM 180,-. Deutlich günstigere Kosten-Leistungs-Relationen werden dagegen für die beiden ADM-Studien mitgeteilt. So beträgt die berichtete Ausschöpfung des Wohlfahrtssurveys 1998, der wie die beiden ALLBUS-Erhebungen von Infratest durchgeführt wurde, 56 Prozent. Bei einem Fallpreis von DM 125,- ist er jedoch wesentlich billiger und auch die Feldzeit ist mit acht Wochen weit niedriger als in den ALLBUS-Studien. Noch mehr zu noch günstigeren Preisen bietet – vermeintlich – der Sozialwissenschaften-Bus 1997, der von GETAS erhoben wurde: Bei einem Fallpreis von DM 115,- und einer Feldzeit von lediglich vier Wochen wird hier eine Ausschöpfung von knapp 70 Prozent ermittelt.

Die Erklärung für dieses paradoxe Ergebnis liegt darin, dass die bei den ADM-Studien dokumentierten Ausschöpfungsquoten sehr unzuverlässig und in der Regel überhöht sind. Die Denkschrift der Deutschen Forschungsgemeinschaft zu „Qualitätskriterien der Umfrageforschung“ – an der auch Vertreter namhafter deutscher Umfrageinstitute mitgewirkt haben – führt zur Frage der Ausschöpfungsberechnung aus: „Die Aussagekraft der errechneten Stichprobenausschöpfung kann unterschiedlich sein. Sie hängt davon ab, wie die verwendeteten Bruttoadressen ausgewählt und dokumentiert werden. In dieser Hinsicht gibt es zwischen verschiedenen Verfahrensmodellen erhebliche Unterschiede. Je „härter“ die Verfahrensregeln für die Bestimmung und Dokumentation des Stichpro-

28 Das weitere Modell – ADM-Stichproben nach dem Adress-Random-Verfahren – bleibt im folgenden außen vor. Es wird in der Praxis eher seltener verwendet und nimmt – was den Ermessensspielraum der Interviewer bei der Auswahl der Zielpersonen sowie die Kontrollmöglichkeiten von Umfrageinstitut und Auftraggeber angeht – eine „mittlere“ Position zwischen ADM-RR-/ADM-SR-Stichproben einerseits, Registerstichproben andererseits ein.

benbrutto sind – was unter Qualitätsgesichtspunkten positiv zu werten wäre -, um so niedriger ist die errechnete Stichprobenausschöpfung." (Kaase 1999: 103f)

Tabelle 7: Ausschöpfung, Feldzeit und Fallpreise in vier Bevölkerungsumfragen*

Studie	Berichtete Ausschöpfungs-quote	Feldzeit	Fallpreis (ca., ohne MWSt.)	Stichproben-verfahren
ALLBUS 2000	49,1%	28 Wo	250 DM	Register
ALLBUS 1996	54,2%	18 Wo	180 DM	Register
Wohlfahrts-survey 1998	56,1%	8 Wo	125 DM	ADM-Random-Route
Sozialwissen-schaften-Bus 1997	68,6%	4 Wo	115 DM**	ADM-Random-Route

* Durchführendes Umfrageinstitut: ALLBUS, Wohlfahrtssurvey: Infratest; Sozialwissenschaften-Bus: GETAS
Fallzahl: ALLBUS 2000 N=3.138, ALLBUS 1996 N=3.518, Wohlfahrtssurvey N=3.042, Sozialwissenschaften-Bus N=2.044
Durchschnittliche Interviewdauer: ALLBUS 2000 ca. 65 Minuten, ALLBUS 1996 ca. 70 Minuten, Wohlfahrtssurvey ca. 50 Minuten, Sozialwissenschaften-Bus ca. 60 Minuten

** Preis errechnet für eine Einschaltung von 55 Minuten Dauer

Bei ADM-Stichproben (zumindest bei dem gängigen Standard-Random- bzw. Random-Route-Modell) wird die Bruttostichprobe im Verlaufe der Feldarbeit durch die Interviewer erstellt. Durch die Vorgabe spezifischer Verfahrensregeln besteht dabei zwar theoretisch kein Spielraum bei der Bestimmung der Zielhaushalte und Zielpersonen. Faktisch aber werden den Interviewern von den Instituten – mehr oder weniger explizit – Kompensationsmöglichkeiten für Ausfälle eingeräumt, um eine schnellere und kostengünstigere Studienbearbeitung zu ermöglichen. Diese Substitution von Ausfällen wird jedoch häufig nicht oder nur eingeschränkt dokumentiert; überhöhte Ausschöpfungsquoten sind die Folge. In dem bereits zitierten DFG-Memorandum heißt es dazu: „Die Standardverfahren der Umfrageinstitute für persönlich-mündliche Befragungen auf random-Basis dürften daher durchweg mit Substitution arbeiten. ... Jeder dokumentierte Substitutionsfall erhöht allerdings die Bruttozahl der eingesetzten Adressen und verringert dementsprechend die errechnete Stichprobenausschöpfung. Es ist daher mit einer gewissen Tendenz zu rechnen, Substitution nicht vollständig zu erfassen und zu dokumentieren." (Kaase 1999: 103)[29]

29 Ein einfaches Rechenexempel verdeutlicht, dass bereits relativ „kleine" Manipulationen die Höhe der berichteten Ausschöpfung merklich beeinflussen können. Angenommen, für einen Inter-

Bei der Registerstichprobe wird die Bruttostichprobe ohne Mitwirken der Interviewer vor der Feldarbeit gezogen, eine unkontrollierte Substitution von Adressen ist nahezu ausgeschlossen und die Ausschöpfung dadurch sehr zuverlässig ermittelbar (zu den Möglichkeiten der Interviewerkontrolle bei Registerstichproben siehe Koch 1995). Die absurde Folge dieser Situation ist, dass die Qualitätsorientierung von Registerstichproben scheinbar doppelt bestraft wird: durch höhere Preise und längere Feldzeiten einerseits, und vermeintlich niedrigere Ausschöpfungsquoten andererseits.[30]

Was kann und sollte in dieser Situation getan werden? Notwendig ist eine offene und sachliche Diskussion von Fragen der Feldarbeit, Datenqualität und Kosten. Diese Aufgabe betrifft sowohl die Umfrageinstitute als auch die potentiellen Auftraggeber von Studien. Einen Ausgangspunkt für die Debatte können die bereits erwähnte DFG-Denkschrift und die „Standards zur Qualitätssicherung in der Markt- und Sozialforschung“ des Arbeitskreises Deutscher Markt- und Sozialforschungsinstitute (1999) bilden. Bislang findet eine entsprechende Diskussion in der Bundesrepublik, aber auch in anderen Ländern, nur in Ansätzen statt (vgl. Schnell 1997: 265f). Zu oft überwiegt auf beiden Seiten das Desinteresse, sei es aus Angst, dass Schwachstellen offengelegt werden, oder aus Furcht vor Kritik, eine Umfrage entspräche nicht geltenden Qualitätsstandards.

Darüber hinaus ist jedoch auch eine stärkere Nachfrage nach anspruchsvollen Verfahrensmodellen unerläßlich, denn in der Umfragepraxis entscheidet letztlich der Markt über die Bedeutung von Qualitätskriterien. „Verfahrensmodelle mit hohen Qualitätsstandards werden auf seiten der anbietenden Institute nur entwickelt und angeboten, wenn es dafür eine Nachfrage gibt, also Auftraggeber, die auf die spezifischen Qualitätskriterien Wert legen und den dafür erforderlichen Aufwand auch zu finanzieren bereit sind.“ (Kaase 1999: 98) In der jüngeren Vergangenheit hat speziell der ALLBUS versucht, im Sinne eines „best practice“ die Stichprobenqualität zu maximieren (vgl. Kaase 1999: 99ff.). Es wäre wünschenswert, wenn weitere Studien auf diesem Weg mitgingen, denn die Imple-

viewer wären in einem Wahlbezirk acht Adressen als Brutto vorgesehen, aus denen er unter Beachtung aller Regeln vier Interviews erzielte. Dann läge seine Ausschöpfung bei 50 Prozent. Erzielte er noch ein weiteres Interview mit einer Person, die nicht zu seinem Brutto zählt, könnte er seine Ausschöpfung bereits auf 62,5 Prozent „verbessern“, wenn er bei der Dokumentation des Bruttos einen der vier Ausfälle unterschlägt.

30 Ein Vergleich der Randverteilungen ausgewählter soziodemographischer Merkmale aus sechs allgemeinen Bevölkerungsumfragen, deren Ausschöpfung zwischen 50 und 80 Prozent variierte, mit Daten des Mikrozensus ergab keinen empirischen Hinweis darauf, dass die Stichprobenverzerrungen der Umfragen mit höherer (berichteter) Ausschöpfung geringer ausfallen (vgl. Koch 1998). Im Gegenteil: Tendenziell am besten schnitt der ALLBUS 1994 ab, der zwar lediglich eine Ausschöpfung von circa 54 Prozent aufwies, aber auf Basis einer Registerstichprobe durchgeführt worden war.

mentierung anspruchsvoller Verfahren der Stichprobenziehung und Feldarbeit wird umso erfolgreicher sein, je eher ein bestimmte „kritische“ Masse von Studien erreicht wird.

Damit Entscheidungen über das für eine Studie am besten geeignete Stichproben- und Felddesign auf einer fundierten Grundlage getroffen werden können, sind die Umfrageinstitute gefordert, stärker als bisher ihre Prozeduren und die damit gewonnenen Erfahrungen offenzulegen und zu dokumentieren. Außerdem ist methodische Forschung notwendig, in der unter kontrollierten Bedingungen unterschiedliche Modelle der Stichprobenbildung und Feldarbeit, die mit ihnen erzielbare Ergebnisqualität und der entstehende Kosten- und Zeitaufwand erforscht werden. Besonderes Augenmerk ist dabei auf die Frage der Verallgemeinerbarkeit der erzielten Resultate zu richten, denn: Zu unterscheiden ist zwischen der prinzipiell erreichbaren Qualität im jeweiligen Verfahrensmodell und der tatsächlich erzielten Qualität bei einer konkreten Umfrage. Mit anderen Worten: Nicht nur vom Verfahrensmodell, sondern auch von seiner Implementation hängt es ab, welche Ergebnisse im konkreten Fall erzielt werden (vgl. Kaase 1999: 107f). Die oben berichteten Ergebnisse vermitteln einen Eindruck davon, welche Unterschiede hinsichtlich Faktoren wie Ausschöpfung, Feldzeit und Kosten auch innerhalb des gleichen Verfahrensmodells auftreten können. Was die eigentliche Datenqualität und deren mögliche Variation innerhalb und zwischen den Verfahrensmodellen angeht, sind weitere Untersuchungen in der Zukunft erforderlich.[31]

Literatur

Alt, Christian/Bien, Walter/Krebs, Dagmar 1991: Wie zuverlässig ist die Verwirklichung von Stichprobenverfahren? Random route versus Einwohnermeldeamtsstichprobe. ZUMA-Nachrichten 28: 65-72.

Arbeitsgemeinschaft ADM-Stichproben und Bureau Wendt 1994: Das ADM-Stichproben-System. Stand 1993. S. 188-202 in: Siegfried Gabler/Jürgen H.P. Hoffmeyer-Zlotnik/Dagmar Krebs (Hrsg.): Gewichtung in der Umfragepraxis. Opladen: Westdeutscher Verlag.

Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V., (Hrsg.) 1999: Standards zur Qualitätssicherung in der Markt- und Sozialforschung. Frankfurt/M.: Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute.

31 Die bislang einzige methodische Vergleichsstudie beispielsweise zwischen einer Register- und einer ADM-Stichprobe in der Bundesrepublik fand 1988 im Rahmen des DJI-Familiensurveys statt. Alt et al. (1991) liefern empirische Indizien dafür, dass die schlechteren Ergebnisse der ADM-Stichprobe – was die Verteilung der Merkmale Geschlecht und Erwerbstätigkeit angeht – auf Manipulationen der Interviewer bei der Auswahl der Befragungspersonen zurückzuführen sind.

Behrens, Kurt/Löffler, Ute 1999: Aufbau des ADM-Stichproben-Systems. S. 69-91 in: ADM Arbeitsgemeinschaft Deutscher Markt- und Sozialforschungsinstitute e.V. und AG.MA Arbeitsgemeinschaft Media-Analyse e.V. (Hrsg.): Stichproben-Verfahren in der Umfrageforschung. Eine Darstellung für die Praxis. Opladen: Leske + Budrich.

Bliesch, Uwe (o.J.): Die Interviewerorganisation in einem Markt- und Sozialforschungsinstitut. München: Infratest.

Blyth, Bill 1998: Current and Future Technology Utilization in European Market Research. S. 563-581 in: Mick P. Couper/Reginald P. Baker/Jelke Bethlehem/Cynthia Z. F. Clark/Jean Martin/William L. Nicholls II/James M. O'Reilly (eds.): Computer Assisted Survey Information Collection. New York: Wiley.

Collins, Martin/Sykes, Wendy/O'Muircheartaigh, Colm 1998: Diffusion of Technological Innovation: Computer Assisted Data Collection in the U.K. S. 23-43 in: Mick P. Couper/Reginald P. Baker/Jelke Bethlehem/Cynthia Z. F. Clark/Jean Martin/William L. Nicholls II/James M. O'Reilly (eds.): Computer Assisted Survey Information Collection. New York: Wiley.

Context 1993: Instituts-Handschriften: Was ist zu tun? Interviewer-Merkmale: Wie sie durchschlagen. Heft 10: 6-10.

Fowler, Floyd J./Mangione, Thomas W., 1990: Standardized Survey Interviewing. Minimizing Interviewer Related Error. Applied Social Research Methods Series, Volume 18. Newbury Park: Sage.

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

Groves, Robert M./Couper, Mick P., 1998: Nonresponse in Household Interview Surveys. New York: Wiley.

Hartmann, Peter H./Schimpl-Neimanns, Bernhard 1992: Sind Sozialstrukturanalysen mit Umfragedaten möglich? Analysen zur Repräsentativität einer Sozialforschungsumfrage. Kölner Zeitschrift für Soziologie und Sozialpsychologie 44: 315-340.

Kaase, Max (Hrsg.) 1999: Qualitätskriterien der Umfrageforschung. Denkschrift der Deutschen Forschungsgemeinschaft. Berlin: Akademie Verlag.

Koch, Achim 1991: Zum Zusammenhang von Interviewermerkmalen und Ausschöpfungsquoten. ZUMA-Nachrichten 28: 41-53.

Koch, Achim 1995: Gefälschte Interviews: Ergebnisse der Interviewerkontrolle beim ALLBUS 1994. ZUMA-Nachrichten 36: 89-105.

Koch, Achim 1997: ADM-Design und Einwohnermelderegisterstichprobe. Stichprobenverfahren bei mündlichen Bevölkerungsumfragen. S. 99-116 in: Siegfried Gabler/Jürgen H.P. Hoffmeyer-Zlotnik (Hrsg.): Stichproben in der Umfragepraxis. Opladen: Westdeutscher Verlag.

Koch, Achim 1998: Wenn „mehr“ nicht gleichbedeutend mit „besser“ ist: Ausschöpfungsquoten und Stichprobenverzerrungen in allgemeinen Bevölkerungsumfragen. ZUMA-Nachrichten 42: 66-90.

Koch, Achim/Wasmer, Martina/Harkness, Janet/Scholz, Elvira 2001: Konzeption und Durchführung der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) 2000. ZUMA-Methodenbericht 2001/05. Mannheim.

Lynn, Peter/Turner, Rachel/Smith, Patten 1998: Assessing the Effects of an Advance Letter for a Personal Interview Survey. Journal of the Market Research Society 40: 265-272.

Niehoff, Renate 1998: Intervieweranforderungen und –auswahl. S. 51-65 in: Statistisches Bundesamt (Hrsg.): Interviewereinsatz und –qualifikation. Spektrum Bundesstatistik, Band 11. Stuttgart: Metzler-Poeschel.

Reinecke, Jost 1998: Regelmäßigkeiten des Interviewerverhaltens. S. 115-137 in: Statistisches Bundesamt (Hrsg.): Interviewereinsatz und –qualifikation. Spektrum Bundesstatistik, Band 11. Stuttgart: Metzler-Poeschel.

Schanz, Volker/Schmidt, Peter 1984: Interviewsituation, Interviewermerkmale und Reaktionen von Befragten im Interview: eine multivariate Analyse. S. 72-113 in: Karl Ulrich Mayer/Peter Schmidt (Hrsg.): Allgemeine Bevölkerungsumfrage der Sozialwissenschaften. Beiträge zu methodischen Problemen des ALLBUS 1980. Frankfurt/M.: Campus.

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

Schnell, Rainer/Kreuter, Frauke 2000: Das DEFECT-Projekt: Sampling-Errors und Nonsampling-Errors in komplexen Bevölkerungsstichproben. ZUMA-Nachrichten 47: 89-102.

Schräpler, Jörg-Peter/Wagner, Gert G., 2001: Das Verhalten von Interviewern – Darstellung und ausgewählte Analysen am Beispiel des „Interviewerpanels“ des Sozio-oekonomischen Panels. Allgemeines Statistisches Archiv 85: 45-66.

Smith, Tom 1995: Trends in Non-Response Rates. International Journal of Public Opinion Research 7: 157-171.

Sodeur, Wolfgang 1997: Interne Kriterien zur Beurteilung von Wahrscheinlichkeitsauswahlen. ZA-Information 41: 58-82.

Statistisches Bundesamt, (Hrsg.) 2000: Datenreport 1999. Zahlen und Fakten über die Bundesrepublik Deutschland. Bonn: Bundeszentrale für politische Bildung.

WER NIMMT TEIL AN PANEL-BEFRAGUNGEN? UNTERSUCHUNG ÜBER DIE BEDINGUNGEN DER ERFOLGREICHEN KONTAKTIERUNG FÜR SOZIALWISSENSCHAFTLICHE UNTERSUCHUNGEN

TATJANA MIKA

Untersucht wird der Einfluss verschiedener Faktoren auf die Bereitschaft sich für sozialwissenschaftliche Untersuchungen befragen zu lassen. Am Beispiel einer Panel-Wiederholungsbefragung wird gezeigt, dass zwar auch Bildung und politische Einstellung, bereits häufig untersuchte Einflussfaktoren, eine gewisse Rolle bei der Teilnahmebereitschaft spielen. Die entscheidenden Faktoren sind aber der positive Verlauf des Interviews und die persönlichen Eigenschaften der befragenden Person. Durch die Auswahl der richtigen Interviewperson lässt sich also eine unerwünschte soziale Selbstauswahl von Befragten verhindern.

The article discusses a number of factors related to individuals' willingness to participate in a social survey. The relevance of socio-demographic characteristics of the interviewees such as social standing, education, and interest in politics are already well established. Using panel study data, the investigation demonstrates that interview-related factors play a decisive role in the decision to participate in a survey. Appropriate selection of interviewers proves to be crucial for optimising continued participation and reducing panel attrition.

1. Einleitung

Die Teilnahme an sozialwissenschaftlichen Befragungen ist zeitaufwendig und anstrengend, vor allem wenn Antworten auf komplexe Fragen erwartet werden. Es ist daher nicht selbstverständlich, dass sich Personen dieser Mühe unterziehen (Sigelmann 1982). Die

Teilnahme wird nicht finanziell honoriert und ist daher ein altruistisches Verhalten, das nicht von allen Mitgliedern der Gesellschaft gleichermaßen erwartet werden kann.[1]

Es gibt einige Hypothesen dazu, welche Typen von Personen sich besonders bereitwillig zu sozialwissenschaftlichen Interviews bereit finden. Es sollen dies zunächst Personen sein, die ein generelles Interesse an dem Gegenstand der jeweiligen Untersuchung haben, so etwa sollen politisch Interessierte bereitwilliger an Umfragen zu politischen Themen teilnehmen (Arzheimer/Klein 2001: 25). Einige Befragte sollen die Teilnahme an einer Umfrage auch als eine Form der politischen Betätigung ansehen, weshalb in Umfragen solche Personen überrepräsentiert sein sollen, die sich aktiv politisch betätigen. Andererseits wird ein Einfluss des sozialen Status auf die Bereitschaft zur Teilnahme vermutet. Es gäbe ein „Mittelschichts-Bias" in der Einstellungsforschung, weil sich mehr Personen mit höherer Bildung, höherem Einkommen und höherer Schichteinstufung zu Befragungen bereit erklären würden (Sigelmann 1982: 347).

Schließlich sollen sich Gründe aus der Interviewsituation ergeben. Deshalb sollen Befragte, die eine positive Beziehung zu dem Interviewer oder der Interviewerin aufgebaut haben, eher bereit sein, sich etwa bei einem Panel, erneut von der gleichen Person befragen zu lassen (Porst/Schneid 1989).

Wenn beide Thesen sich bisher empirisch validieren ließen, dann stellt sich die Frage nach ihrer Gewichtung. Die folgende Untersuchung zeigt anhand der Daten des Panels des *International Social Justice Project* (ISJP) von 2000 und 2001, dass den Interviewereffekten der größere Einfluss auf die mögliche Verzerrung der Stichprobe durch Selbstselektion zugemessen werden sollte. Dies ist insofern ein positives Ergebnis, als die Umstände des Interviews - etwa durch die Auswahl der Interviewer/innen - beeinflusst werden können.

2. Bereitschaft zur Teilnahme an der Panel-Befragung

Die Gründe für die Teilnahme an sozialwissenschaftlichen Befragungen lassen sich besonders gut anhand von Panel-Umfragen untersuchen, weil hier eine Reihe von Informationen über diejenigen vorliegen, die einmal ein Interview gegeben haben, sich aber der erneuten Befragung verweigern. Zur Teilnahme an der Panelbefragung werden in der Regel diejenigen ausgewählt, die sich am Ende der ersten Befragung zu einer weiteren

1 Bei einigen Untersuchungen werden den Befragten Geschenke, u.a. auch Geldgeschenke, gegeben. Der Wert dieser Geschenke erreicht aber in keinem Fall das Niveau einer am Zeitaufwand der Befragten orientierten Bezahlung. Der Anschein einer Bezahlung soll sogar gerade vermieden werden (Dillmann, D. A. 1978: *Mail and Telephone Surveys. The Total design Method*. New York: Wiley).

Teilnahme freiwillig bereit erklären (Porst/Briel 1995: 11). Diese werden dann für die Wiederholungsbefragung erneut kontaktiert und, wenn diese Kontaktierung erfolgreich verläuft, ein weiteres Mal interviewt.

Das *International Social Justice Project* (ISJP) ließ 2000 in Ost- und Westdeutschland eine repräsentative Befragung in allen Bundesländern zu Einstellungen zu Fragen der sozialen Gerechtigkeit durchführen. Für diese Befragung wurden 3215 Interviews in ganz Deutschland durchgeführt. An diese Untersuchung schloss sich eine aktuell noch laufende Panel Studie an, in deren Rahmen ein Teil der im Jahr 2000 Interviewten im Abstand von jeweils einem Jahr noch zwei weitere Male befragt werden, also in den Jahren 2001 und 2002. Für diese Panelstudie wurden die Bundesländer Baden Württemberg, Berlin und Sachsen ausgewählt. Für die folgende Untersuchung werden daher nur die Befragten aus diesen Bundesländern herangezogen. Die Grundgesamtheit der im Jahr 2000 in diesen Bundesländern Befragten war 966 Personen.

Von den Befragten der ersten Welle des ISJP war 2000 eine große Anzahl bereit, sich noch einmal befragen zu lassen. In Ostdeutschland war dabei, wie in Tabelle 1 zu sehen ist, die Bereitschaft generell größer als in Westdeutschland. Es kam dann noch zu einer Reduzierung des Umfangs der Stichprobe, weil einige der im Jahr 2000 Befragten nicht mehr aufgefunden werden konnten, verstorben waren oder die Befragung verweigerten. Diese Reduktion war vom Umfang her in Ost- und Westdeutschland ähnlich groß, aber die Gründe waren in Ostdeutschland häufiger Umzug und nicht angetroffen werden, im Westen häufiger Verweigerung.

Tabelle 1: Bereitschaft zur zweiten Befragung und tatsächliche Teilnahme

	Ost		**West**		**Gesamt**	
Bereitschaft zur Teilnahme an der Panel Befragung						
	%	N	%	N	%	N
Ja	79,71	444	65,53	268	73,71	712
Nein	20,29	113	34,47	141	26,29	254
Gesamt	100	557	100	409	100	**966**
Teilnahme an der Panel Befragung						
	%	N	%	N	%	N
Teilgenommen	60,32	336	55,50	227	58,28	563
Nicht teilgenommen	39,68	221	44,50	182	41,72	403
Gesamt	100	557	100	409	100	**966**

3. Unterschiede in der Sozialstruktur bei Selbstselektion

Es wird häufig vermutet, dass Mitglieder bestimmter sozialer Gruppen sich leichter für sozialwissenschaftliche Befragungen motivieren lassen. Hierzu soll die Mittelschicht gehören.

Zunächst interessieren uns die Unterschiede zwischen denjenigen, die sich 2000 zu einem weiteren Interview bereit erklären und solchen, die sich nicht noch ein weiteres Mal befragen lassen wollen, denn die Bereitschaft an der Befragung teilzunehmen, ist der wichtigste Filter für die erneute Befragung. Diese Ergebnisse stellen wir dann den Unterschieden bei den gleichen Merkmalen in der tatsächlich befragten Panel Stichprobe gegenüber. Hier zeigt sich die Wirkung der Selbstselektion zum Zeitpunkt der Kontaktierung durch das Befragungsinstitut. Wir werden sehen, dass die Veränderungen der Stichprobe in Ostdeutschland und Westdeutschland spezifisch unterschiedlich sind und auf unterschiedliche Gründe für die Verweigerung der Teilnahme an der Befragung hinweisen.

a) Sozialstruktur

In Tabelle 2 sehen wir ein verblüffendes Phänomen. Nach der Selbstauswahl der Befragten anhand der Bereitschaft droht eine Verzerrung der Stichprobe in Ostdeutschland zugunsten der Mittelschicht und der höher Gebildeten, wie sie nach der Hypothese des „Mittelschichtbias" zu erwarten ist. Unter den tatsächlich Befragten findet sich diese Verschiebung aber nicht mehr, stattdessen taucht eine Verschiebung in Westdeutschland auf, die nach der Befragung der Bereitschaft gar nicht zu erwarten gewesen wäre.

Tabelle 2: Unterschiede in der Sozialstruktur (Mittelwerte)

	Ost		**West**		**Gesamt**	
Bereitschaft zur Teilnahme an der Panel Befragung						
	Nein	**Ja**	**Nein**	**Ja**	**Nein**	**Ja**
Bildung	3,65	4,16**	3,77	3,60	3.71	3.95*
Schicht	4,41	5,04**	5,77	5,55	5.16	5.23
Teilnahme an der Panel Befragung						
	Nicht teilg.	**Teilg.**	**Nicht teilg.**	**Teilg.**	**Nicht teilg.**	**Teilg.**
Bildung	4,12	4,01	3,64	3,67	3,90	3,88
Schicht	4,95	4,88	5,93	5,38**	5.39	5.08**

* $p_t<.05$; ** $p_t<.01$; Bildung: Casmin level 1 (less than general)- level 7 (higher tertiary vocational training); Schicht, Selbsteinstufung auf einer Skala der sozialen Schichtung von 1 = „unten" bis 10 = „oben"

Die Mechanismen der Selbstselektion werden also in Westdeutschland bei der Kontaktierung für die Zweitbefragung wirksam, noch nicht bei der Erklärung der Bereitwilligkeit im Jahr zuvor. Es lässt sich hier vermuten, dass die Entscheidung zugunsten des Interviews in Westdeutschland spontaner getroffen wird als in Ostdeutschland. Hierfür spricht auch, dass die Rate der Verweigerung nach erfolgreicher Kontaktierung in Westdeutschland erheblich höher war als in Ostdeutschland.[2]

b) Politische Einstellung

Eine weitere Hypothese vermutet, dass ein Interesse am Inhalt der Umfrage die (wiederholte) Teilnahme befördert (Arzheimer/Klein 2001). Politisch Interessierte äußern sich, gemäß dieser These, bereitwilliger in Umfragen zu politischen Themen. Es wird darüber hinaus vermutet, dass diejenigen, die an Panelbefragungen teilnehmen, dies zum Teil als eine Form der politischen Meinungsäußerung ansehen. Soziale Gerechtigkeit, der Schwerpunkt der ISJP Befragung, ist ein Thema, dass eine starke Nähe zu einer Reihe von politischen Themen hat und vor allem für politisch weiter links stehende von stärkerem Interesse sein könnte.[3] Deshalb untersuchen wir die Unterschiede in der politischen Einstellung anhand der Selbsteinstufung als politisch eher links oder eher rechts stehend.

Daher wollen wir nun untersuchen, ob sich die Befragten in ihren politischen Einstellungen von den aus der Befragung Ausgeschiedenen unterscheiden, so dass es zu einer Verzerrung in Bezug auf die politischen Einstellungen kam. Dabei unterscheiden wir zunächst die Gruppe der sich zur Wiederholung bereit Erklärenden und der dann tatsächlich Befragten. Die politische Einstellung wird hier durch eine 10-stufige Skala gemessen, wobei die kleinen Werte für eine linkere und große Werte für eine rechtere Einstufung stehen.

Wir sehen, dass es bei der Auswahl anhand der Bereitschaft zur erneuten Teilnahme noch keine Unterschiede innerhalb der Teilpopulationen gibt. Es kommt nur zu einer Verschiebung bei der gesamtdeutschen Stichprobe nach links, weil die Ostdeutschen sich in stärkerem Maße zu einer Wiederholungsbefragung bereit erklärten und insgesamt eine signifikant linkere politische Einstellung haben.[4]

---

2 In Baden Württemberg verweigerten 23,5 % der Kontaktierten das Interview, in Sachsen 11,9% und in Berlin (Ost und West) nur 7%. Quelle: USUMA Methodenbericht.

3 Dies ergibt sich daraus, dass vor allem linkere Parteien mit dem Thema „soziale Gerechtigkeit" assoziiert werden.

4 Der Unterschied in der politischen Einstellung ist auf dem 1% Niveau signifikant.

Tabelle 3: Unterschiede in der politischen Einstellung (Mittelwerte)

	Ost		West		Gesamt	
Bereitschaft zur Teilnahme an der Panel Befragung						
	Nein	**Ja**	**Nein**	**Ja**	**Nein**	**Ja**
Politische Einstellung	5,12	4,81	5,49	5,50	5.32	5.07*
Teilnahme an der Panel Befragung						
	Nicht teilg.	**Teilg.**	**Nicht teilg.**	**Teilg.**	**Nicht teilg.**	**Teilg.**
Politische Einstellung	4,98	4,80	5,78	5,27**	5,34	4,99**

* p_t<.05; ** p_t<.01; Politische Einstellung, Einstufung auf 10 stufiger Skala 1 = sehr links, 10 = sehr rechts

Bei der Realisierung der Interviews kommt es dann allerdings in Westdeutschland zu einer deutlichen Linksverschiebung. Diese Beobachtung entspricht der Verschiebung in Westdeutschland beim Merkmal „Schichteinstufung“, die wir oben gesehen haben.

4. Verhalten im Interview

Nach der These von Hoag (Hoag 1981) zeichnen sich Personen, die mehrfach an Panel Befragungen teilnehmen, dadurch aus, dass ihre Antwortbereitschaft von den Interviewern besser eingeschätzt wird. Auch halten die Interviewer ihre Antworten für zuverlässiger. Wir untersuchen auch diese These auf ihre Bedeutung bei den sich bereit Erklärenden und den tatsächlich Befragten.

Wir finden einen deutlichen Unterschied bei der Einschätzung der Antwortbereitschaft durch den Interviewer/die Interviewerin, wie sich aus Tabelle 4 ergibt.

Diejenigen, die sich zu einer weiteren Befragung bereit erklärten, tendierten klar zu einem Wert gegen 1 („gut“), diejenigen, die nicht zur Befragung bereit waren, tendieren zur 2 („mittelmäßig“). Auch bei der Zuverlässigkeit der Antworten gibt es Anhaltspunkte für die Richtigkeit dieser These. Hierbei findet sich kein signifikanter Ost-West Unterschied, aber jeweils ein signifikanter Unterschied zwischen denjenigen, die zu einer erneuten Befragung bereit waren, und denen, die sich einer solchen Befragung verweigerten. Dieser Unterschied verschwindet bei den Ostdeutschen, wenn die tatsächlich Teilnehmenden untersucht werden, bleibt aber in Westdeutschland bestehen.

Tabelle 4: Unterschiede beim Antwortverhalten (Mittelwerte)

	Ost		West		Gesamt	
Bereitschaft zur Teilnahme an der Panel Befragung						
	Nein	**Ja**	**Nein**	**Ja**	**Nein**	**Ja**
Antwortverhalten	1,74	1,13**	1,89	1,24**	1,82	1,17**
Zuverlässigkeit	1,15	1,08*	1,19	1,07**	1,17	1,07**
Teilnahme an der Panel Befragung						
	Nicht teilg.	**Teilg.**	**Nicht teilg.**	**Teilg.**	**Nicht teilg.**	**Teilg.**
Antwortverhalten	1,43	1,22**	1,61	1,41*	1,51	1,30**
Zuverlässigkeit	1,13	1,08	1,18	1,07*	1,15	1,08**

* p_t<.05; ** p_t<.01; zu Antwortbereitschaft: 1 = gut, 2 = mittelmäßig, 3=anfangs gut/später schlechter, 4=anfangs schlecht/später besser, 5=schlecht; zur Zuverlässigkeit: 1 = insgesamt zuverlässig, 2 = insgesamt weniger zuverlässig/bei einigen Fragen weniger zuverlässig

5. Interviewereffekte

Einen wesentlichen Einfluss auf den erfolgreichen Verlauf einer *face to face* Befragung hat die Person der Interviewerin/des Interviewers (Scheuch, 1967: 156/157). Die Befragten reagieren auf die interviewende Person, in der Regel ohne dass dieser ihre Wirkung bewusst ist (Bortz/Döring 2002: 246). Die eingesetzten Interviewer/innen waren gebeten worden ihr Alter, ihr Bildungsniveau und ihre Geschlecht anzugeben. Es zeigt sich, dass die eingesetzten Interviewer/innen in Ost- und Westdeutschland ein sehr unterschiedliches Sozialprofil haben. Im Osten interviewten mehr Männer und höher Gebildete als im Westen. Auch das Alter lag deutlich höher.

Tabelle 5: Sozialprofil Interviewer/innen

	Ost	West	Gesamt
Sozialprofil der Interviewer/innen			
Geschlecht weiblich	28%	45%	35%
Altersdurchschnitt	58 Jahre	50 Jahre	54 Jahre
Hochschulabsolvent/innen	71%	30%	54 %

Wir werden nun untersuchen, welchen Einfluss diese Merkmale der Interviewer/innen, neben den bisher untersuchten Merkmalen der Interviewten und des Interviewverlaufs, auf die Wahrscheinlichkeit der Teilnahme an einer Wiederholungsbefragung haben.

Die bisher gefundenen Unterschiede wurde in einer multinominal-logistischen Regression auf ihre relative Bedeutung für die tatsächliche Teilnahme an einer Wiederholungsbefragung getestet. Hierfür wurden drei Gruppen unterschieden. Die erste Gruppe enthält diejenigen, die sich zur Panelteilnahme bereit erklärt hatten, und auch tatsächlich befragt werden konnten („Freiwillige"). Sie konstituieren in der Regression die Referenzgruppe für die beiden anderen Gruppen. Die Vergleichsgruppen bilden diejenigen, die sich zur Teilnahme bereit erklärt hatten, aber dann doch nicht erfolgreich kontaktiert werden konnten[5] („Drop-outs"), und diejenigen, die nicht eine weiteres Mal befragt werden wollten („Verweigerer").

Die Untersuchung soll zeigen, ob sich die erfolgreiche Gewinnung für ein weiteres Interview stärker durch die Charakteristika der Interviewten oder eher durch die Merkmale des Interviews und der Befragungsperson erklären lässt.

Der beste Erklärungsfaktor für die erneute Teilnahme an der Befragung ist die positive verlaufene Kommunikation im ersten Interview, die als gut eingeschätzte Antwortbereitschaft der Interviewten.

Der nächste große Erklärungsfaktor sind die Charakteristika der Interviewer/innen. Es zeigt sich, dass vor allem das Alter und Bildungsniveau der Interviewer/innen einen großen Einfluss auf die Teilnahme an einer weiteren Befragung haben. Die Verweigerung eines weiteren Interviews lässt sich zu einem großen Anteil durch ein „junges" Alter und zu hohe Bildung der Interviewer/innen erklären. Der Bildungseffekt tritt dabei vor allem in Ostdeutschland auf, wo das Bildungsniveau der Interviewer auch erheblich höher war als in Westdeutschland.[6] Die Personen, die ein weiteres Interview im Osten verweigert haben, wurden erheblich häufiger von Hochschulabsolvent/innen befragt, was offenbar nicht von allen Befragten positiv aufgenommen wurde. Die Westdeutschen reagieren auf diesen Punkt weniger, wie der Interaktionseffekt im dritten Modell zeigt. Das Geschlecht der Interviewer/innen spielt dagegen weder in Ost- noch in Westdeutschland eine Rolle.[7]

---

5 Nur eine sehr geringe Anzahl von Befragten war aus der Untersuchung ausgeschieden, weil sie verstorben waren (zwischen 0 und 1,7%). Die übrigen wurden nicht mehr angetroffen (unbekannt verzogen u. ä.) oder verweigerten die Teilnahme.

6 Es wurde auch getestet, ob die Befragten positiv auf Interviewer/innen mit einem ihrem eigenen ähnlichen Bildungsniveau reagieren. Dies ist jedoch nicht der Fall. Alle Befragten reagieren positiv auf Interviewer/innen mit niedrigerem Bildungsniveau.

7 Es gibt auch keinen Effekt derart, dass ein entgegengesetztes Geschlecht der Intervier/innen zu größerer Bereitschaft der Teilnahme führt.

Tabelle 6: Multinominal- logistische Regression zur Modellierung des Teilnahmeverhaltens

	Modell I: Struktur und Bewertung der Antwortbereitschaft		**Modell II: Interviewermerkmale**		**Modell III: Interaktionseffekt**	
	Drop-outs	*Verweigerer*	*Drop-outs*	*Verweigerer*	*Drop-outs*	*Verweigerer*
Westdeutschland	0.051	0.606	-0.090	0.492	-0.112	1.210
	(0.30)	(3.23)**	(0.48)	(2.32)*	(0.45)	(3.63)**
Schicht	0.063	-0.002	0.040	-0.006	0.041	-0.005
	(1.20)	(0.03)	(0.74)	(0.11)	(0.77)	(0.08)
Bildung	0.148	0.063	0.137	0.013	0.136	0.032
	(2.54)*	(0.96)	(2.27)*	(0.19)	(2.25)*	(0.46)
Politische Einstellung	0.101	0.003	0.099	0.014	0.098	0.008
	(2.03)*	(0.06)	(1.96)	(0.24)	(1.94)	(0.14)
Antwortbereitschaft	0.709	1.657	0.724	1.754	0.721	1.745
	(4.08)**	(10.02)**	(4.12)**	(10.26)**	(4.11)**	(10.20)**
Interviewer: Alter			-0.037	-0.058	-0.037	-0.061
			(4.84)**	(6.63)**	(4.75)**	(6.87)**
Interviewer: Hochschulabschluss			0.346	0.741	0.321	1.448
			(1.81)	(3.39)**	(1.36)	(4.33)**
Interviewer: Weiblich			0.084	-0.397	0.084	-0.381
			(0.46)	(1.90)	(0.46)	(1.82)
Interaktionseffekt: Hochschulabschluss (Interviewer) * Westdeutschland					0.050	-1.298
					(0.14)	(2.92)**
Konstante	-2.969	-4.642	-0.744	-1.074	-0.923	-0.547
	(5.07)**	(7.22)**	(0.86)	(1.12)	(1.29)	(0.69)
R^2	0.1011		0.1320		0.1374	
N	951		951		951	

* $p_t<.05$; ** $p_t<.01$; Multinominal- logistische Regression (maximum likelihood), Referenzgruppe: Teilnahme

Die politische Einstellung oder die soziale Stellung der Interviewten verliert dagegen an Bedeutung, wenn für die anderen Faktoren kontrolliert wird. Dies zeigt, dass in einem *face to face* Interview auch Personen mit geringerer Bildung und geringerem Einkommen für ein Interview zu gewinnen sind, was bei schriftlichen Befragungen im Allgemeinen schlechter gelingt (Schützenmeister, 2002, S.151). Allerdings bleibt der Unterschied

zwischen Ost- und Westdeutschland bestehen. Gerade die geringere Befragungsbereitschaft in Westdeutschland konnte hier nicht so gut aufgeklärt werden.

6. Fazit

Anhand einer Panel Befragung zum Themenbereich soziale Gerechtigkeit wurde untersucht, welche Befragten der ersten Welle sich zu einer Wiederholungsbefragung bereit erklärten und auch an dieser teilnahmen. Obwohl die Bereitschaft zu einer erneuten Teilnahme generell hoch war, gab es in Ostdeutschland deutliche Verzerrungen bei der erklärten Bereitschaft, sich ein weiteres Mal befragen zu lassen. Diese drohende Verzerrung der Stichprobe durch die stärkere Bereitschaft zur Wiederholungsbefragung von höher gebildeten und sich in der Schicht höher einstufenden Befragten, konnte durch die Ausfälle bei der Kontaktierung wieder ausgeglichen werden.

In Westdeutschland ergibt sich die Veränderung der Stichprobe dagegen nicht aus der Antwort auf die Frage nach der Bereitschaft zur Teilnahme, sondern tritt bei der Realisierung der Interviews ein. Wir müssen leider beobachten, dass die Ausfälle in Westdeutschland Auswirkungen auf die soziodemografische Zusammensetzung und die politische Tendenz der Stichprobe haben.

Der stärkste Indikator für eine erfolgreiche weitere Kontaktierung ist allerdings die erfolgreiche Kommunikation zwischen Interviewer/in und Befragtem im ersten Interview. Wenn dieses positiv verläuft, dann gelingt häufig auch die Kontaktierung zur Wiederholungsbefragung. Dabei spielt die Person des/r Interviewers/in eine entscheidende Rolle. Diejenigen, die nicht zu einer weiteren Befragung bereit waren, sind häufiger von einer weniger alten und höher gebildeten Person befragt worden. Das relativ jüngere Alter der Interviewpersonen spielt dabei in Ost- und Westdeutschland, die Hochschulbildung hauptsächlich in Ostdeutschland eine Rolle bei der Verweigerung einer weiteren Befragung. Durch eine an diese Ergebnisse angepasste Auswahl der Interviewpersonen könnte die Ausschöpfungsrate mit Sicherheit erhöht werden.

Korrespondenzadresse

Tatjana Mika
ISJP – Humboldt Universität zu Berlin
Institut für Sozialwissenschaften
Unter den Linden 6, 10999 Berlin
tatjana.mika@rz.hu-berlin.de

Literatur

Arzheimer, K./M. Klein 2001: Die Wirkung materieller Incentives auf den Rücklauf einer schriftlichen Panelbefragung. ZA-Information 43: 6-31.

Bortz, J./N. Döring 2002: Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Berlin: Springer.

Dillmann, D. A., 1978: Mail and Telephone Surveys. The Total design Method. New York: Wiley.

Hoag, W., 1981: Realisierte Stichproben bei Panels: Eine vergleichende Analyse. ZUMA Nachrichten 9: 6-18.

Porst, R./Ch. von Briel 1995. Wären Sie vielleicht bereit, sich gegebenenfalls noch einmal befragen zu lassen? Oder: Gründe für die Teilnahme an Panel-Befragungen. Mannheim: ZUMA.

Porst, R./M. Schneid 1989: Ausfälle bei der Panel-Befragung. In: Planung und Analyse 1/89: 8-13.

Scheuch, E. K., 1967: Das Interview in der Sozialforschung. In: König, René (Hrsg.), Handbuch der empirischen Sozialforschung, I. Band. Stuttgart: Ferdinand Enke, 136-196.

Schützenmeister, F., 2002: Die Bereitschaft, sich wieder befragen zu lassen, in postalischen Erhebungen. Zeitschrift für Soziologie 31: 138-154.

Sigelmann, L., 1982: The uncooperative Interviewee. Quality and Quantity 16: 345-353.

Auswirkungen des Erhebungsverfahrens bei Jugendbefragungen zu ‚Heiklen' Themen – Schulbasierte Schriftliche Befragung und Haushaltsbasierte Mündliche Befragung im Vergleich[1]

Dietrich Oberwittler und Thomas Naplava

Schriftliche Befragungen im Klassenverband sind heute in der empirischen Jugendforschung weit verbreitet, ohne dass die methodischen Aspekte dieser Erhebungsform als ausreichend untersucht gelten können. In einer Methodenstudie haben wir daher eine haushaltsbasierte mündliche Befragung und eine schulbasierte schriftliche Befragung bei identischer Grundgesamtheit und identischem Erhebungsinstrument durchgeführt. Die Ausschöpfungsrate der schulbasierten Befragung liegt erheblich höher als die der haushaltsbasierten Befragung, und insbesondere Personen aus unteren sozialen Schichten werden besser erreicht. Am Beispiel der selbstberichteten Delinquenz ergibt der Vergleich der Befragungsergebnisse erhebliche Unterschiede sowohl in den Prävalenzen als auch in den Korrelationen mit anderen Variablen. In der haushaltsbasierten Befragung liegen die Prävalenzraten der selbstberichteten Delinquenz ungefähr 20 bis 50 Prozent niedriger als in der schulbasierten Befragung. Während wir die unterschiedlichen Prävalenzraten eher mit den Selektionseffekten der Stichprobenverfahren erklären, deuten die unterschiedlichen Korrelationsmuster möglicherweise auf Kontexteffekte der jeweiligen Erhebungssituationen hin.

1 Der Beitrag entstand im Rahmen des Forschungsprojekts „Soziale Probleme und Jugenddelinquenz im sozialökologischen Kontext", das von der Deutschen Forschungsgemeinschaft finanziell gefördert wird (Ob 134/3-1 und –2). Wir danken Tom Blank und Tilman Köllisch sowie dem anonymen Gutachter für Kommentare zu den Entwurfsfassungen.

Using paper-and-pencil surveys in classrooms is common practice in empirical adolescence research. However, the methodological impacts of this mode of data collection are only partially understood. To gain more insight into these issues, two surveys were conducted, sampling respondents from the same population and using the same questionnaire but employing different modes of data collection: a household-based survey with face-to-face interviews and a school-based survey with paper-and-pencil interviews. The response rate of the school-based survey was much higher, and more people from the lower social strata are included. Response differences can be noted across modes. A comparison of the rates of self-reported delinquency, for example, revealed large differences both in the prevalence and the correlations of delinquency. Prevalence rates in the household-based survey are about 20 to 50 per cent lower than in the school-based survey. While the different prevalences may best be explained by a selection effect of data collection methods, the different correlation patterns would seem to point more to mode effects of interview situations.

1. Einleitung

1.1 Problemstellung und Zielsetzung

Empirische Untersuchungen zu jugendsoziologischen Themen werden heute häufig mit Hilfe von Schulbefragungen – schriftliche Befragungen von SchülerInnen im Klassenverband während des Unterrichts – durchgeführt (z.B. Currie et al. 2000; Merkens 1999; Zinnecker et al. 2002). Insbesondere Schulbefragungen zum Thema Jugenddelinquenz und -gewalt haben in den vergangenen Jahren eine bemerkenswerte Konjunktur erlebt (Heitmeyer et al. 1995; Holtappels et al. 1997, Mansel/Hurrelmann 1998; Tillmann et al. 1999; Wetzels et al. 2001), während entsprechende Studien in den 1980er Jahren häufig haushaltsbasierte Stichproben und face-to-face Interviews nutzten (Albrecht et al. 1988; Schumann et al. 1987; Villmow/Stephan 1983). Eine Schulbefragung liegt bei Studien nahe, für die Schule expliziter Bestandteil der Fragestellung ist (z.B. ‚Gewalt in der Schule'), und zeichnet sich darüber hinaus durch forschungspragmatische Vorteile aus. Schulbefragungen sind vergleichsweise billig und bieten daher die Chance, in kurzer Zeit sehr große Stichproben zu realisieren. Dass diese Chance auch genutzt wird, illustriert die Zahl von 28.368 befragten SchülerInnen, die sich durch die Addition der Stichproben von lediglich fünf im Jahr 2001 veröffentlichten Studien zum Thema Jugenddelinquenz in Deutschland ergibt (Fuchs et al. 2001; Mansel 2001; Oberwittler et al. 2001; Sturzbecher 2001; Wetzels et al. 2001). Daneben werden in der Jugendforschung auch weiterhin haushaltsbasierte Stichprobenverfahren und mündliche Interviews durchgeführt, etwa bei den bekannten ‚Shell-Jugendstudien' (Deutsche Shell 2000, 2002).

Die Auswirkungen dieser unterschiedlichen Erhebungsverfahren, die Vergleichbarkeit der Ergebnisse, die Vor- und Nachteile von Schulbefragungen gegenüber anderen Erhebungsverfahren bleiben jedoch in der wissenschaftlichen Diskussion unterbelichtet. In der einschlägigen Literatur wird die Schulbefragung, bei der es sich um eine Sonderform der schriftlichen Befragung handelt, nur am Rande oder gar nicht erwähnt, da sie bei der erwachsenen „Normalbevölkerung" naturgemäß nur selten eingesetzt werden kann (Deutsche Forschungsgemeinschaft 1999; Diekmann 1998; Friedrichs 1990; Schnell et al. 1999). Es liegen nur sehr wenige Methodenstudien vor, die die spezifischen Eigenschaften und Probleme von schriftlichen Befragungen im Gruppenkontext durch Vergleiche mit anderen Erhebungsverfahren oder durch Variationen der Erhebungssituation zu klären versuchen (Kerkvliet 1994; Kreuzer et al. 1992). Einerseits gilt das mündliche Interview aus methodischer Sicht vielfach immer noch als die bevorzugte Erhebungsmethode, während die schriftliche Befragung als zwar kostengünstige, jedoch qualitativ unterlegene Alternative angesehen wird (vgl. Klein/Porst 2000; Reuband 2001); andererseits bieten schriftliche Befragungstechniken bei der Erforschung ‚heikler' Themen wegen der größeren Vertraulichkeit der Erhebungssituation spezifische Vorteile (Dillman 1983: 375; Reuband/Blasius 1996). Haushaltsbasierte mündliche Befragungen und schriftliche Schulbefragungen unterscheiden sich jedoch nicht nur hinsichtlich der Erhebungssituation, sondern auch hinsichtlich des Feldzugangs, der Stichprobenziehung und der Ausschöpfungsquoten voneinander. Es kann angenommen werden, dass alle Aspekte Auswirkungen auf die Befragungsergebnisse haben.

Ziel dieses Beitrages ist es daher, anhand eines Vergleiches der beiden maßgeblichen Erhebungsmethoden der quantitativen Jugendforschung – einer schriftlichen Schulbefragung und einer haushaltsbasierten mündlichen Befragung – bei derselben Grundgesamtheit und demselben Erhebungsinstrument Unterschiede in den Befragungsergebnissen hinsichtlich von Häufigkeiten und Zusammenhängen zu untersuchen; dieser Vergleich erfolgt am Beispiel des Themas ‚selbstberichtete Delinquenz'. Im folgenden sollen zunächst theoretische Überlegungen erstens zu den Unterschieden im Feldzugang und in den Ausschöpfungsraten und zweitens zu den Unterschieden in der Erhebungssituation und dem Antwortverhalten angestellt werden, bevor wir anschließend empirische Ergebnisse unserer methodischen Vergleichsstudie präsentieren. Da bei dem Vergleich von haushaltsbasierter mündlicher Befragung und schriftlicher Schulbefragung sowohl der Feldzugang als auch die Befragungssituation gleichzeitig variiert werden, ist eine klare Trennung der unterschiedlichen Effekte jedoch nicht möglich. Es handelt sich also um einen – unseres Wissens jedoch den ersten – vorrangig deskriptiven Vergleich der beiden in der Jugendforschung häufigsten methodischen Designs; unsere abschließenden Erklärungsansätze der vorgefundenen Unterschiede bleiben notwendigerweise spekulativ.

1.2 Theoretische Überlegungen zu Feldzugang und Ausschöpfungsraten

Als bedeutsamer Unterschied zwischen haushalts- und schulbasierten Befragungen von Jugendlichen fallen die erheblich höheren Ausschöpfungsquoten der Schulbefragungen ins Auge. Diese liegen häufig zwischen 80 Prozent und 90 Prozent der bereinigten Bruttostichproben, die in der Regel als die Gesamtzahl der SchülerInnen in den ausgewählten Schulklassen definiert werden.[2] So hohe Ausschöpfungsquoten sind in der Praxis haushaltsbasierter Befragungen zumindest dieser Altersgruppe illusorisch (Schnell 1997). Der Unterschied in den Ausschöpfungsquoten ist plausibel, da durch den Feldzugang über die Institution Schule einige Schwellen entfallen, die den Zugang zu den Zielpersonen über die privaten Haushalte charakterisieren. Kosten und Nutzen der Teilnahme bzw. Verweigerung verteilen sich in den beiden Erhebungsformen gegensätzlich. Während bei haushaltsbasierten Befragungen aus der Sicht der Befragten durch die Teilnahme Kosten anfallen, z.B. durch das Opfern freier Zeit, entstehen bei Schulbefragungen eher Kosten durch die Verweigerung, da diese gegen die Erwartungen der Klasse und der Lehrer durchgesetzt werden muss. Der Ausfall des normalen Unterrichts während einer Schulbefragung wird von den meisten Jugendlichen als Nutzen bewertet.

Niedrige Ausschöpfungsquoten stellen dann ein Problem dar, wenn zu erwarten ist, dass die Teilnahme an einer Befragung von Merkmalen abhängig ist, die auch mit den Zielvariablen der Untersuchung zusammenhängen. Generell könnten schichtspezifische Stichprobenverzerrungen relevant sein; bekanntlich verweigern Zielpersonen mit niedrigem Sozial- und speziell Bildungsstatus häufiger die Teilnahme an Befragungen als andere Personen (Schnell 1997: 205). In welchem Ausmaß dieser Effekt auch bei einer jugendlichen Zielpopulation auftritt, und wie groß die Unterschiede zwischen haushaltsbasierter und Schulbefragung ausfallen, ist jedoch unklar, denn auch bei Schulbefragungen ist mit einer schultypspezifischen Nonresponse vor allem an Haupt- und Sonderschulen zu rechnen.

Systematische Ausfälle von Personen mit bestimmen (extremen) Ausprägungen können gerade bei der Erforschung von Jugenddelinquenz eine bedeutsame Rolle spielen. Das gilt zum Beispiel für die – wenn auch sehr kleinen – Gruppen der obdachlosen und strafgefangenen Jugendlichen, die praktisch bei keiner Bevölkerungsstichprobe berücksichtigt werden; offensichtlich ist dies auch bei Schulbefragungen zum Thema Schulschwänzen der Fall, da Schwänzer in der Schule schlechter erreichbar sind als Nicht-Schwänzer. Eine systematisch geringere Teilnahmebereitschaft delinquenter Jugendlicher an Befragungen –

---

2 Dabei werden Ausfälle auf Schulebene als qualitätsneutrale Ausfälle angesehen, da die Gründe für eine Verweigerung auf der Institutionenebene überwiegend andere sind als auf der Ebene der Individuen, und die Entscheidung für oder gegen eine Befragungsteilnahme i.d.R. nicht durch die Individuen beeinflusst werden kann; z.B. Wetzels et al. 2001: 71

unabhängig von ihrem sozialen Status – könnte plausibel mit der Kontrolltheorie erklärt werden, die schwache Bindungen an die Institutionen der konventionellen Gesellschaft als Ursache für delinquentes Verhalten annimmt (Hirschi 1969). Gilt dies zunächst unabhängig vom Untersuchungsthema, kommt bei Befragungen zur Delinquenz hinzu, dass delinquente Jugendliche negative Konsequenzen ihrer Angaben zur eigenen Delinquenz befürchten könnten. Daher ist es bei entsprechenden Befragungen üblich, den eigentlichen Untersuchungsgegenstand bei der Kontaktaufnahme zu verschleiern.

Diese Überlegungen führen zu den folgenden Hypothesen über die Auswirkungen der besseren Ausschöpfungsraten bei Schulbefragungen im Vergleich zu haushaltsbasierten Befragungen:

- Schulbefragungen erreichen eine bessere Abbildung von Jugendlichen mit niedrigem Sozialstatus.
- Schulbefragungen erreichen eine bessere Abbildung von delinquenten Jugendlichen.

Während die erste Hypothese relativ leicht zu testen ist, wird die Überprüfung der zweiten Hypothese durch den mit dem Stichprobenverfahren konfundierten Effekt der Befragungstechnik auf das Antwortverhalten erschwert.

1.3 Theoretische Überlegungen zu Erhebungssituation und Antwortverhalten

Bei der Beantwortung sogenannter heikler Fragen spielt die Erhebungssituation eine wichtige Rolle. Als heikel werden gemeinhin Fragen bezeichnet, deren wahrheitsgemäße Beantwortung zu einer besonders großen Spannung mit der Tendenz der sozialen Erwünschtheit führen kann (Diekmann 1998: 383). Hierzu zählen insbesondere verschiedene Formen abweichenden Verhaltens wie zum Beispiel bestimmte Sexualpraktiken oder strafbares Verhalten. Durch den Effekt der sozialen Erwünschtheit könnte die Hemmung, besonders stigmatisierte Handlungen (z.B. Gewaltdelikte) zuzugeben, größer sein als bei anderen, weniger stigmatisierten Verhaltensweisen (z.B. Haschischkonsum). Dabei ist jedoch zu bedenken, dass die wenigsten Normen Allgemeingültigkeit besitzen, sondern dass in der Bevölkerung unterschiedliche und teils entgegengesetzte Bewertungsmaßstäbe für soziale Erwünschtheit existieren; dies gilt besonders stark für jugendliche (Sub-)Kulturen. So dürfte etwa Drogenkonsum innerhalb einer Gruppe von Techno-Fans als weniger ‚heikel' empfunden werden als in einem katholischen Priesterseminar. Bei der Selbstdarstellung hängt das Ausmaß der sozialen Erwünschtheit zusätzlich von der subjektiv wahrgenommenen normativen Erwartungshaltung der Adressaten ab: Während Jugendliche gegenüber erwachsenen Autoritätspersonen in der Regel eine soziale Erwünschtheit in der Richtung der Drogenabstinenz verspüren werden, mag dieser Druck in bestimmten

Gleichaltrigengruppen mit entsprechenden normativen Überzeugungen sogar in die entgegengesetzte Richtung des Drogenkonsums gehen.

Die Situation bei einer mündlichen Befragung in der privaten Wohnung des Befragten ist durch die persönliche Nähe von Interviewer und Befragtem und durch eine relativ geringe Anonymität geprägt. Das gilt auch, wenn die heiklen Fragen – wie zum Beispiel nach der eigenen Delinquenz – vom Hauptfragebogen abgetrennt und vom Befragten selbst ausgefüllt werden, oder wenn CASI-Techniken zum Einsatz kommen. Denn zwar entfällt die Notwendigkeit, dem Interviewer heikle Sachverhalte „ins Gesicht" sagen zu müssen, jedoch bleibt eine nachträgliche personenbezogene Auswertung der Antworten durch das Befragungsinstitut (bis hin zur Weitergabe der Informationen an die Polizei) zumindest technisch möglich.

Die Art und Richtung von Interviewereffekten wird allgemein durch die Nähe bzw. Distanz sozialer Merkmale zwischen Befragten und Interviewern bestimmt. Bei Fragen zu abweichendem Verhalten sind hinsichtlich großer Altersspannen zwischen Befragten und Interviewern Interviewereffekte zu erwarten, da Personen verschiedenen Alters Normverletzungen in unterschiedlichem Maß bewerten. Reuband (1985) z.B. konnte zeigen, dass Jugendliche, die von deutlich älteren Interviewern befragt wurden, weniger häufig Drogenkonsum, aber genauso häufig Alkoholkonsum angaben im Vergleich zu Jugendlichen, die von nur geringfügig älteren Interviewern befragt wurden. Allgemein wirkt sich die vom Befragten wahrgenommene Nähe der Werte und Normen der Interviewer zu den eigenen Werten positiv auf die Antwortbereitschaft aus, so dass Merkmale wie Bildung und Lebensstil der Interviewer für die Befragungssituation von besonderer Bedeutung sind (Schnell u.a. 1999). Hindelang u.a. (1981) konnten allerdings zeigen, dass bei ihrer Studie von vier verschiedenen, auf die Stichproben zufällig verteilten Interviewbedingungen keine zu unterschiedlichen Prävalenzraten der selbstberichteten Delinquenz geführt hat. Die Prävalenzraten variierten weder zwischen der Form der Erhebung (face-to-face gegenüber schriftlicher Befragung), noch zwischen dem Ausmaß an zugesicherter Anonymität.

Schulbefragungen unterscheiden sich von haushaltsbasierten Befragungen durch eine ungleich größere Anonymität der Befragungssituation. Der Interviewer tritt einem Kollektiv ihm unbekannter SchülerInnen gegenüber und bleibt während der schriftlichen Befragung in einer relativ großen Distanz zu den einzelnen Befragten. Unter diesen Bedingungen ist den Befragten plausibel zu vermitteln, dass ihre Angaben vertraulich behandelt und nicht persönlich zugeordnet werden können (Dillman 1983: 375).

Sind also bei schriftlichen Schulbefragungen Interviewereffekte eher unwahrscheinlich, so können Effekte durch die spezifische Situation des Gruppenkontextes, in dem die Befragung stattfindet, auftreten. Die Erfahrungen bei Schulbefragungen zeigen, dass das

Ausfüllen der Fragebögen in den normalen Klassenräumen nicht völlig unbeobachtet von den Sitznachbarn möglich ist. Angaben können vom Nachbarn übernommen werden, und die Möglichkeit des Einblicks in den Fragebogen durch den Nachbarn kann wiederum auch Hemmungen im Antwortverhalten hervorrufen (Beebe et al. 1998). In Klassen, in denen viele Jugendliche Delinquenz bejahende Orientierungen haben, wären Kontexteffekte in der Richtung denkbar, dass Befragte Delikte unwahrheitsgemäß bzw. in übertriebenem Ausmaß angeben, um sich damit bei ihren Mitschülern zu profilieren. Hier wäre ein umgekehrter Effekt der sozialen Erwünschtheit (im Sinne einer jugendlichen Subkultur) wirksam.

Insgesamt sprechen diese Überlegungen für die folgende Hypothese:

- Die Erhebungsituation bei schulbasierten schriftlichen Befragungen beeinflusst das Antwortverhalten der Befragten in Richtung größerer Häufigkeiten der selbstberichteten Delinquenz im Vergleich zur Erhebungssituation bei haushaltsbasierten mündlichen Befragungen.

Das bedeutet, dass sowohl Ausschöpfung bzw. Nonresponse als auch das Antwortverhalten für unterschiedliche Ergebnisse in den beiden Befragungsformen verantwortlich sein könnten. Eine eindeutige Zuordnung von Unterschieden zu dem einen oder anderen Aspekt ist daher bei der vorliegenden Studie nicht möglich und muss spekulativ bleiben.

2. Daten und Ergebnisse

2.1 Stichproben

Für den empirischen Vergleich der beiden Erhebungsmethoden stehen zwei Jugendbefragungen zur Verfügung, die im Rahmen des Projekts „Soziale Probleme und Jugenddelinquenz im sozialökologischen Kontext“ in Freiburg i.Br. im Herbst 1999 bzw. im Frühjahr 2000 durchgeführt wurden. Den aus beiden Befragungen gewonnenen Samples für den Methodenvergleich liegt mit der männlichen jugendlichen Wohnbevölkerung Freiburgs im Alter von 15 und 16 Jahren, die allgemeinbildende Schulen besuchen, dieselbe Grundgesamtheit zugrunde. Die Fragebögen, die bei beiden Studien eingesetzt wurden, beinhalteten grundsätzlich identische Instrumente in fast identischer Reihenfolge, nur die Auswahl und Anzahl der Fragen in den Fragebögen wiesen einige Unterschiede auf.[3] Beide Erhebungsinstrumente umfassen ca. 230 Einzelvariablen. Für die Erfassung der selbstberichteten Delinquenz wurde in beiden Studien dieselbe, aus 16 Items bestehende

3 Diese Unterschiede bestehen überwiegend im hinteren, auf die zentrale Skala der selbstberichteten Delinquenz nachfolgenden Teil des Fragebogens.

Skala eingesetzt (Oberwittler et al. 2002), so dass Auswirkungen unterschiedlicher Erhebungsinstrumente auf die Angaben zur selbstberichteten Delinquenz ausgeschlossen werden können.

Die Schulbefragung fand im Zeitraum Oktober bis Dezember 1999 in 91 Schulklassen der 8. bis 10. Jahrgangsstufen an 24 staatlichen und privaten Schulen im Stadtgebiet Freiburg sowie zeitgleich in Köln mit einer wesentlich größeren Stichprobe statt. Die Ausschöpfungsrate auf der Institutionenebene der Schulen beträgt in Freiburg 83 Prozent; bei den verweigernden Schulen handelte es sich um eine Hauptschule, zwei Realschulen und zwei Gymnasien. Auf der Ebene der SchülerInnen in den zufällig ausgewählten Klassen innerhalb der teilnehmenden Schulen beträgt die Ausschöpfungsquote 86 Prozent (90 Prozent an Gymnasien und Realschulen, 75 Prozent an Sonderschulen), die realisierte Stichprobe umfasst 1886 männliche und weibliche Befragte. Für die Teilnahme an der Befragung war entsprechend den schulrechtlichen Rahmenbedinungen die schriftliche Einverständniserklärung eines Erziehungsberechtigten erforderlich; unter den teilnehmenden SchülerInnen wurden je Klasse eine Freikarte für Kino oder Sportveranstaltungen verlost. Die Befragung während ein bis zwei Schulstunden wurde von geschulten InterviewerInnen in den normalen Klassenzimmern durchgeführt. Als explizite Verweigerer sind 3,3 Prozent der Zielpersonen in der Bruttostichprobe bekannt; bei den übrigen 10,7 Prozent Nicht-Teilnehmern, die am Befragungstag nicht im Klassenraum anwesend waren, muss offen bleiben, ob ihre Abwesenheit mit der angekündigten Befragung im Zusammenhang steht.

Die haushaltsbasierte mündliche Befragung fand im Mai und Juni 2000 statt und wurde durch das Freiburger Institut für angewandte Sozialwissenschaft e.V. an einer Zufallsstichprobe ausschließlich männlicher Jugendlicher im Alter zwischen 14 und 17 Jahren aus dem Freiburger Einwohnermelderegister durchgeführt. Jugendliche, die bereits an der einige Monate zuvor durchgeführten Schulbefragung teilgenommen hatten, wurden nicht befragt. Zielpersonen mit nicht-deutscher Staatsangehörigkeit wurde überproportional häufig gezogen; ihr Anteil beträgt in der bereinigten Bruttostichprobe (N= 934) 28,7 Prozent gegenüber 12,9 Prozent in der Grundgesamtheit.[4] Durch einen Fehler des Meldeamtes bei der Adressenziehung wurden in der Gruppe der Deutschen deutlich weniger 14jährige Personen ausgewählt, so dass diese Altersgruppe in der Bruttostichprobe mehrheitlich nicht-deutsch und insgesamt seltener vertreten ist als die anderen Altersgruppen. Wegen dieser Verzerrung mussten die 14jährigen Befragten für die hier berichteten Auswertungen aus beiden Stichproben ausgeschlossen werden. Die überproportionale Ziehung von nicht-deutschen Personen in der haushaltsbasierten Stichprobe wird durch einen

4 Dies bezieht sich auf die Jugendlichen im Alter zwischen 14 bis 16 Jahren beiderlei Geschlechts.

entsprechenden Gewichtungsfaktor berücksichtigt.[5] Die face-to-face-Interviews wurden von geschulten InterviewerInnen in den Wohnungen der Befragten durchgeführt und dauerten durchschnittlich 47 Minuten.

Die realisierte Stichprobe der mündlichen Befragung umfasst 498 Befragte, was einer Ausschöpfungsquote von 53,3 Prozent entspricht. Als wichtigste Gründe der Nicht-Teilnahme sind die Verweigerung durch die Jugendlichen (30,6 Prozent) bzw. der Erziehungsberechtigten (3,0 Prozent), die Nichterreichbarkeit nach fünf Kontaktversuchen (8,2 Prozent) und die Nichtauffindbarkeit der Adressen (3,2 Prozent) zu nennen. Aus rechtlichen Überlegungen wurde in dem Kontaktanschreiben darauf hingewiesen, dass auch delinquentes Verhalten von Jugendlichen zu den Themen der Studie gehört.

Für die Vergleichbarkeit beider Befragungen wurden jeweils übereinstimmende Teilstichproben definiert (siehe Übersicht in Tabelle 1). Von der Stichprobe der Schulbefragung wurden zunächst alle weiblichen Jugendlichen und alle Jugendlichen mit einem Wohnsitz außerhalb Freiburgs ausgeschlossen. Von der Stichprobe der mündlichen Befragung wurden alle Befragten ausgeschlossen, die nicht mehr allgemeinbildende Schulen besuchen, da angenommen werden kann, dass sie sich in einer deutlich veränderten sozialen Lebenslage befinden. Sodann wurden die Stichproben hinsichtlich der Altersstruktur angepasst. Zum einen wurden zusätzlich zu den 14jährigen auch die 13jährigen Befragten aus der Stichprobe der Schulbefragung ausgeschlossen, da diese Altersgruppe kein Bestandteil der Stichprobe der mündlichen Befragung ist; zum anderen wurden auch die über 16jährigen ausgeschlossen, da diese Altersgruppe sich überwiegend nicht mehr in der 10. Jahrgangsstufe befindet; diejenigen über 16jährigen, die sich in die Stichprobe der Schulbefragung befinden, haben zu 85 Prozent bereits eine oder mehrere Klassen wiederholt, so dass von einer inhaltlich bedeutsamen Stichprobenverzerrung auszugehen ist. Aus diesem Grunde wurden aus beiden Stichproben nur die Altersgruppe der 15- und 16jährigen berücksichtigt. Es bleiben für den Vergleich 385 Befragte in der schulbasierten Stichprobe und 324 Befragte in der haushaltsbasierten Stichprobe. Die Ausschöpfungsquote der haushaltsbasierten Stichprobe liegt auch nach Ausschluss der 14- und 17jährigen Personen bei 53,3 Prozent.[6]

---

5 In der Grundgesamtheit (Bevölkerungszahlen liegen vor für die Altersgruppe 15 bis 16 Jahre beiderlei Geschlechts) befinden sich 3037 deutsche und 440 nicht-deutsche Personen, in der bereinigten Bruttostichprobe 481 deutsche und 159 nicht-deutsche Personen. Daraus wurde ein Gewichtungsfaktor von 0.4383 errechnet.

6 Hierbei werden auch 25 Befragte mitgezählt, die keine allgemeinbildende Schule besuchen, da diese Information für die Bruttostichprobe nicht bekannt ist.

Tabelle 1: Übersicht über die Definition der Vergleichs-Stichproben

	mündliche Befragung →	Vergleichssample	Schul- ← Befragung
Alter	14 bis 17 Jahre[a]	15 bis 16 Jahre	ca. 13 bis 16 Jahre (8. bis 10. Jahrgangsstufe)
Nationalität	Alle	alle	alle
Geschlecht	nur männlich	nur männlich	männlich / weiblich
Zielpopulation	Jugendliche in Privathaushalten in Freiburg	Schüler an allgemeinbildenden Schulen, in Privathaushalten in Freiburg lebend	Schüler an allgemeinbildenden Schulen in Freiburg
Stichproben-Umfang	498 →	324+385 = 709	← 1886

a) Adressenziehung für 14-Jährige fehlerhaft, daher im Vergleichssample ausgeschlossen

2.2 Ausschöpfungsraten und sozio-demographische Merkmale

Zunächst soll die Frage untersucht werden, inwieweit die unterschiedlichen Stichprobenverfahren und insbesondere die daraus resultierenden unterschiedlichen Ausschöpfungsquoten zu Unterschieden in der sozio-demographischen Zusammensetzung der beiden Stichproben im Vergleich zueinander und zur Grundgesamtheit geführt haben. Anhand von vier Indikatoren (Staatsangehörigkeit bzw. ethnische Herkunft, Bildungsniveau der Befragten und ihrer Eltern sowie elterlicher Berufsstatus) kann festgestellt werden, dass Personen mit niedrigem und mittlerem sozialen Status in der haushaltsbasierten Stichprobe schlechter repräsentiert werden als in der schulbasierten Stichprobe.

Da die (deutsche vs. nicht-deutsche) Staatsangehörigkeit der Zielpersonen in der Bruttostichprobe der mündlichen Befragung bekannt ist, können getrennte Ausschöpfungsquoten für deutsche und nicht-deutsche Jugendliche berechnet werden. Diese liegen bei 57,6 Prozent für deutsche und 40,9 Prozent für nicht-deutsche Zielpersonen. Entsprechend liegt der Anteil der nicht-deutschen Befragten in der realisierten Stichprobe (nach Gewichtung) mit 9,2 Prozent niedriger als in der Grundgesamtheit (12,9 Prozent). Neben einer höheren Verweigerungsrate der nicht-deutschen Jugendlichen (34,0 Prozent, deutsch 30,8 Prozent) sind hierfür vor allem eine höhere Nicht-Auffindbarkeit der Adresse (8,2 Prozent bzw. 1,9 Prozent) und eine höhere Nicht-Erreichbarkeit nach fünf Kontaktversuchen (10,1 Prozent bzw. 6,4 Prozent) verantwortlich.

In der schulbasierten Stichprobe ist keine deutlich geringere Ausschöpfung nicht-deutscher Zielpersonen feststellbar. Hier ist ein Vergleich mit der Bruttostichprobe allerdings

nur auf der Aggregatebene der Klassen möglich, für die der Anteil aller SchülerInnen ohne deutsche Staatsangehörigkeit bekannt ist. Da die Angaben von Jugendlichen zu ihrer Staatsangehörigkeit erfahrungsgemäß nicht sehr zuverlässig sind, wurde in beiden Befragungen stattdessen nach den Herkunftsländern der Eltern gefragt. Dem Anteil der SchülerInnen ohne deutsche Staatsbürgerschaft von 12,7 Prozent in der schulbasierten Bruttostichprobe steht ein etwas höherer Anteil von befragten SchülerInnen nicht-deutscher ethnischer Herkunft in der Nettostichprobe von 12,9 Prozent gegenüber;[7] auf der Basis der aus der haushaltsbasierten Stichprobe bekannten Abweichung zwischen Herkunftsland der Eltern und rechtlicher Staatsangehörigkeit entspräche das einem Anteil von Befragten ohne deutsche Staatsangehörigkeit von 12,3 Prozent. Weitere Analysen auf der Aggregatebene der Schulklassen lassen den Schluss zu, dass in den Klassen mit niedriger Ausschöpfungsquote deutsche und nicht-deutsche SchülerInnen gleichermaßen der Befragung ferngeblieben sind; da jedoch die Ausschöpfungsquoten in Klassen mit einem hohen Anteil nicht-deutscher SchülerInnen generell niedriger liegen (siehe unten), kommt es auch bei der schulbasierten Stichprobe zu einer geringfügig niedrigeren Ausschöpfung der nicht-deutschen Bevölkerung.

Das Bildungsniveau der befragten Jugendlichen wird durch den besuchten Schultyp gemessen. Die Verteilungen der beiden Stichproben können mit der amtlichen Schulstatistik verglichen werden, die jedoch auch die weiblichen Schülerinnen und die SchülerInnen mit einem Wohnort außerhalb Freiburgs umfasst (Tabelle 2). Da diese beiden Gruppen häufiger ein Gymnasium oder eine Waldorfschule und seltener eine Sonder- oder Hauptschule besuchen, wäre durch die Einschränkung der beiden Stichproben auf männliche Schüler mit Wohnsitz in Freiburg eine leichte Verschiebung der Anteile von den höheren zu den niedrigeren Bildungsniveaus zu erwarten. Tendenziell, insbesondere bei der haushaltsbasierten Stichprobe, ist jedoch das Gegenteil der Fall. Die Verteilung der schulbasierten Stichprobe stimmt recht genau mit der Verteilung der Grundgesamtheit überein – mit Ausnahme der Gesamtschule[8]; bei der haushaltsbasierten Stichprobe fällt vor allem der mit 1 Prozent sehr niedrige Anteil der Sonderschüler ins Auge. Dies legt den Schluss nahe, dass Sonderschüler bei einem haushaltsbasierten Stichprobenverfahren sehr schwer zu erreichen sind. Allerdings muss auch an die Möglichkeit gedacht werden, dass die Befragten im mündlichen Interview wegen des ausgeprägten sozialen Stigmas dieser

---

7 Diese Angaben beziehen sich auf die gesamte schulbasierte Stichprobe (N=1884), da auf der Aggregatebene nicht nach Geschlecht und Wohnort differenziert werden kann.

8 Der deutlich höhere Anteil von Gesamtschülern in der schulbasierten Stichprobe ergibt sich aus dem Umstand, dass es in Freiburg nur eine Gesamtschule gibt, und diese in die Schulstichprobe aufgenommen wurde, während es mehr Schulen der anderen Schultypen gibt als die ausgewählten. Daher hatten Gesamtschüler eine größere Chance als andere Schüler, in die Stichprobe zu gelangen.

Schulform bewusst unwahre Angaben gemacht haben. Wenn man annimmt, dass sie in diesem Falle häufig ‚Hauptschule' angegeben haben, dann läge der summierte Anteil der Sonder- und Hauptschüler in der haushaltsbasierten Stichprobe mit 15,8 Prozent dennoch deutlich unterhalb des entsprechenden Anteils in der schulbasierten Stichprobe (22,1 Prozent) und der Grundgesamtheit (25,3 Prozent). Insgesamt sprechen diese Verteilungen dafür, dass Zielpersonen mit niedrigem Bildungsniveau in der schulbasierten etwas schlechter und in der haushaltsbasierten Stichprobe deutlich schlechter erreicht werden als Zielpersonen mit mittlerem oder höherem Bildungsniveau.

Tabelle 2: Bildungsstatus (besuchte Schulform) der Befragten im Vergleich zur Grundgesamtheit

	Grundgesamtheit[a] n=16269	Schulbefragung		Mündliche Befragung	
	%	abs.	%	abs.[b]	%[c]
Sonderschule	9,0	25	6,5	6	1,0
Hauptschule	16,3	60	15,6	54	14,8
Realschule	22,6	85	22,1	87	26,6
Gesamtschule	7,4	56	14,5	23	6,9
Gymnasium, Waldorfschule	44,7	159	41,3	154	50,7
Summe	100,0	385	100,0	324	100,0

a Männliche und weibliche SchülerInnen mit Wohnort innner- und außerhalb Freiburgs; gewichtet zur Korrektur der Schülerzahlen in den Jahrgangsstufen 11-13 an Gymnasien und Gesamtschule
b ungewichtet
c gewichtet für die überproportionale Ziehung nicht-deutscher Befragter

Für die Messung des elterlichen Status stehen die – naturgemäß nicht unproblematischen – Angaben der befragten Jugendlichen zu den Bildungsabschlüssen und zur Berufstätigkeit ihrer Eltern zur Verfügung, wobei jeweils der höchste Wert von Vater oder Mutter gezählt wird. Da die Verteilung des elterlichen Bildungs- und Berufsstatus in der Grundgesamtheit unbekannt ist, sind nur Vergleiche der beiden Stichproben untereinander möglich.

Tabelle 3: Höchster Bildungsstatus der Eltern nach Erhebungsform (in %)

	ohne Abschluss	Haupt-schule	Real-schule	(Fach)-Abitur	(Fach)Hoch-schule
alle[a] (X^2=15,9, p=0.003, tau-b=0.08, p=0.03)					
Schulbefragung	2,1	18,2	29,2	27,1	23,4
mündliche Befragung	1,0	20,0	22,1	21,0	35,9
Deutsche[b] (X^2=25,4, p<0.001, tau-b=0.08, p=0.06)					
Schulbefragung	0,7	9,5	36,3	28,8	24,7
mündliche Befragung	0,4	16,4	23,0	20,9	39,3
nicht-Deutsche[b] (X^2=4,2, p=0.375, tau-b=0.04, p=0.56)					
Schulbefragung	6,8	47,7	5,7	20,5	19,3
mündliche Befragung	5,0	41,3	15,0	20,0	19,0

a gewichtet für die überproportionale Ziehung nicht-deutscher Befragter
b Deutsche: Herkunftsland von Vater oder Mutter Deutschland; Nicht-Deutsche: Herkunftsland von Vater und Mutter nicht Deutschland

Tabelle 4: Höchstes Berufsprestige der Eltern nach Erhebungsform (in %)

	sehr niedrig[a]	niedrig[a]	hoch[a]	sehr hoch[a]
alle Befragten[b] (X^2=9,90, p=0.019, tau-b=0.10, p=0.004				
Schulbefragung	18,5	31,8	32,3	17,4
Mündl. Befragung	15,3	24,4	34,1	26,1
Deutsche[c] (X^2=5,89, p=0.117, tau-b=0.07, p=0.077)				
Schulbefragung	16,5	33,5	30,4	19,6
Mündl. Befragung	16,4	26,5	28,8	28,3
Nicht-Deutsche[c] (X^2=0.7, p=0.995, tau-b=0.01, p=.904)				
Schulbefragung	40,0	30,0	20,0	10,0
Mündl. Befragung	38,7	30,7	21,3	9,3

a Rekodierung der Magnitude-Prestige Skala: sehr niedrig 20-53,5 Punkte, niedrig 53,6-87,1 Punkte, hoch 87,2-120,7 Punkte, sehr hoch 120,8-187,8 Punkte; 50 Fälle wurden direkt den vier Klassen zugeordnet
b gewichtet für die überproportionale Ziehung nicht-deutscher Befragter
c Deutsche: Herkunftsland von Vater oder Mutter Deutschland; Nicht-Deutsche: Herkunftsland von Vater und Mutter nicht Deutschland

Beim elterlichen Bildungsstatus sind Abweichungen zwischen beiden Stichproben vor allem in den mittleren und oberen Statusgruppen erkennbar (Tabelle 3).[9] Mittlere Reife oder (Fach-)Abitur als höchsten elterlichen Bildungsabschluss geben 56,3 Prozent der Befragten in der schulbasierten Stichprobe gegenüber nur 43,1 Prozent in der haushaltsbasierten Stichprobe an; dagegen beträgt der Anteil der Befragten mit Eltern mit (Fach-)Hochschulabschluss in der haushaltsbasierten Stichprobe 35,9 Prozent gegenüber nur 23,4 Prozent in der schulbasierten Stichprobe. Insgesamt besteht eine schwache, signifikante Tendenz zu höheren Bildungsniveaus in der haushaltsbasierten Stichprobe (tau-b=0.08). Auffällig ist weiterhin, dass diese Tendenz nur bei den deutschen Befragten besteht (tau-b=0.08, knapp sign.), während die Verteilung der Stichproben bei den nicht-deutschen Befragten fast identisch ist (tau-b=0.04, p=0.56).

Die Ermittlung des Berufsprestige basiert auf der ISCO-Vercodung der offenen Berufsabfrage und anschließender Zuordnung zu den Werten der Magnitude-Prestigeskala von Wegener (1988), die in vier ordinale Klassen rekodiert wurden (Tabelle 4).[10] Es besteht eine deutliche Tendenz zu höheren Berufsprestigewerten in der haushaltsbasierten Stichprobe (tau-b=0.10). Wie beim elterlichen Bildungsstatus besteht der deutlichste Unterschied in der höchsten Prestigeklasse (26,1 Prozent in der haushaltsbasierten Stichprobe, 17,4 Prozent in der schulbasierten Stichprobe); doch auch bei den beiden unteren Prestigeklassen summiert sich die Differenz – hier zugunsten der schulbasierten Stichprobe – auf über 10 Prozentpunkte. Auch beim Berufsprestige bestehen – allerdings nicht mehr signifikante – Unterschiede nur in der Gruppe der deutschen (tau-b=0.07), nicht in der Gruppe der nicht-deutschen Befragten (tau-b=0.01).

Auf der Basis der hier verwendeten Statusindikatoren zeigt sich, dass Zielpersonen aus niedrigeren sozialen Schichten in einem schulbasierten Stichprobenverfahren besser erreicht werden können als in einem haushaltsbasierten Stichprobenverfahren; bei letzterem ist eine stärkere schichtspezifische Selektion wirksam, für die am ehesten eine grö-

---

9 18 Prozent der Befragten der Schulbefragung und 7 Prozent der Befragten der mündlichen Befragung konnten keine Angabe zum Bildungsabschluss ihrer Eltern machen. Weitere Analyen ergaben, dass das durchschnittliche elterliche Berufsprestige der deutschen Befragten ohne gültige Werte der Gruppe der deutschen Befragten mit elterlichem Bildungsabschluss „Realschule“ entsprach, und das der nicht-deutschen Befragten ohne gültige Werte der Gruppe der nicht-deutschen Befragten mit elterlichem Bildungsabschluss „Hauptschule“. Daher wurde den Befragten mit fehlenden Werten ersatzweise diese entsprechenden Werte zugewiesen.

10 Bei 13 Prozent der Befragten der Schulbefragung und 10 Prozent der Befragten der mündlichen Befragung konnte kein ISCO-Wert vergeben werden. Bei jeweils 7% der Befragten wurden die ungenauen Berufsnennungen ersatzweise durch zwei Rater direkt der 4stufigen Ordinalskala zugeordnet; es bleiben 6 Prozent fehlende Werte bei der Schulbefragung und 2 Prozent fehlende Werte bei der mündlichen Befragung übrig.

ßere soziale Distanz zwischen Interviewer bzw. Forschungsinstitut und Zielpersonen und eine generell niedrigere Kooperationsbereitschaft dieser Bevölkerungsgruppen verantwortlich sein dürfte. Angesichts einer sehr hohen Ausschöpfungsquote von 86 Prozent besteht jedenfalls eine vergleichbare schichtspezifische Verzerrung bei dem schulbasierten Stichprobenverfahren nicht.

Dies wird auch deutlich, wenn man abschließend den Zusammenhang zwischen sozialem Status und Ausschöpfungsquoten auf der Aggregatebene betrachtet. Für die jeweiligen Aggregate der beiden Stichprobenverfahren – Schulen bei der schulbasierten Stichprobe, Stadtteile bei der haushaltsbasierten Stichprobe – steht der amtliche Anteil der nicht-deutschen Personen als gemeinsamer Schichtindikator bereit.[11] Die beiden Streudiagramme zeigen, dass die Ausschöpfungsquoten bei dem haushaltsbasierten Stichprobenverfahren nicht nur durchschnittlich niedriger liegen, sondern auch wesentlich stärker vom Sozialstatus der Stadtteile abhängig sind (Abbildungen 1 und 2). Die Neigung der Regressionsgerade der schulbezogenen Ausschöpfungsquoten ist demgegenüber erheblich flacher; auch in den Schulen mit sehr niedrigem Sozialstatus der Schüler liegt die Ausschöpfungsquote noch bei ca. 60 Prozent,[12] während sie in den Stadtteilen mit niedrigem Sozialstatus bis auf 20 Prozent abfällt. Zusammen mit den Ergebnissen auf Individualdatenebene kann daraus geschlossen werden, dass der soziale Status bei haushaltsbasierten Stichprobenverfahren eine größere selektive Wirkung entfaltet als bei schulbasierten Stichprobenverfahren. Die Höhe der Koeffizienten von maximal .10 zeigt jedoch an, dass dieser Selektionseffekt keine dramatischen Ausmaße erreicht.

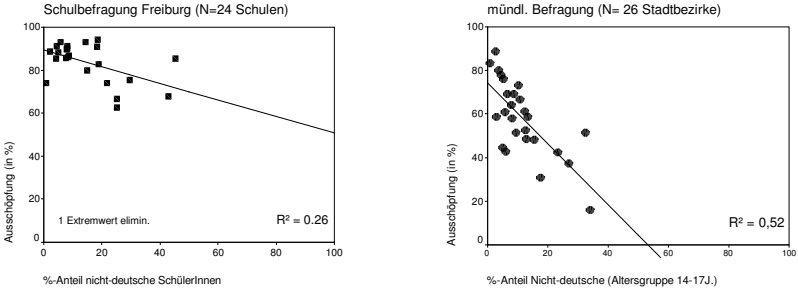
Bemerkenswert erscheint uns, dass dieses Ergebnis nur für die deutsche Teilpopulation gilt, während die Art des Stichprobenverfahrens offenbar keine Auswirkung auf die soziale Zusammensetzung der nicht-deutschen Befragten hat. Obwohl die Ausschöpfung der nicht-deutschen Zielpersonen in der haushaltsbasierten Stichprobe deutlich schlechter ist als in der schulbasierten Stichprobe, finden sich dort nicht weniger Befragte mit niedrigem Sozialstatus. Wie ist dieses auch für uns überraschende Ergebnis zu erklären? Zunächst ist zu berücksichtigen, dass sich die nicht-deutsche Bevölkerung in der Grundgesamtheit überwiegend auf die unteren Statusgruppen konzentriert und die Varianz in den Schichtvariablen daher nicht so groß ist wie in der deutschen Bevölkerung, die sich breiter auf das gesamte Spektrum der sozialen Schichtung verteilt. Dies vermindert den statistischen Spielraum für Verzerrungseffekte. Dennoch ist anzunehmen, dass bei nicht-

11 Der Anteil der nicht-deutschen Bevölkerung korreliert auf der Ebene der Freiburger Stadtteile mit r=0.77 mit der Sozialhilfequote.

12 Nach Schulformen unterschieden beträgt die Ausschöpfungsquote an Gymnasien 90,0 Prozent, an Realschulen 90,3 Prozent, an Hauptschulen 77,7 Prozent und an Sonderschulen 75,0 Prozent.

deutschen Zielpersonen noch andere, nicht gemessene Faktoren für die Frage der Teilnahme bzw. Nichtteilnahme an Befragungen eine Rolle spielen.

Abbildung 1 und 2: Streudiagramme der Ausschöpfungsraten mit dem Anteil der Nicht-Deutschen in den Aggregaten Schule bzw. Stadtbezirk



2.2.1 Zusätzliche Informationen über Nonresponse bei Schulbefragungen

Trotz hoher Ausschöpfungsquoten bleibt auch bei schulbasierten Stichproben die Frage relevant, ob bedeutsame Verzerrungseffekte durch Nonresponse auftreten. Da über die Grundgesamtheit sehr wenige Informationen bekannt sind, kann diese Frage nur durch Befragungen von ‚konvertierten' Nichtteilnehmern (Schnell 1997: 152) beantwortet werden. Dazu haben wir im Rahmen der parallel durchgeführten Schulbefragung in Köln zwei verschiedene Gruppen von Nicht-Teilnehmern befragt: Erstens 27 HauptschülerInnen an fünf Hauptschulen, die am Befragungstag abwesend gewesen waren, aber einige Tage später an einem Nachholtermin befragt wurden; zweitens 38 Jugendlichen, die trotz Schulpflicht dauerhaft den Schulbesuch verweigern, jedoch in sog. „Schulschwänzer-Projekten" sozialpädagogisch betreut werden, wobei diese Befragten nicht notwendigerweise zur Bruttostichprobe gehörten.

Durch die Durchführung eines zweiten Befragungstermins – bei Schulbefragungen wegen des organisatorischen Aufwandes selten oder nie praktiziert – konnte die Ausschöpfungsquote an den fünf betroffenen Hauptschulen von 74 Prozent auf 81 Prozent angehoben werden. Die folgenden Vergleiche beziehen sich ausschließlich die Befragten der fünf Hauptschulen, die am „Haupttermin" bzw. am „Nachholtermin" befragt wurden (Tabelle 5). Aufgrund der geringen Fallzahlen gibt es fast keine signifikanten Unterschiede zwischen beiden Gruppen, jedoch weisen mehrere Variablen auf einen etwas höheren Sozialstatus der am Nachholtermin befragten SchülerInnen hin, bei denen es sich überwiegend

um Mädchen handelt. Durch die nachträgliche Befragung abwesender SchülerInnen kann demnach eine angenommene schichtspezifische Verzerrung der Stichprobe bei Schulbefragungen nicht behoben werden. Erwartungsgemäß ist der Anteil der Schulschwänzer unter den am Nachholtermin Befragten höher, und auch eine geringere Schulmotivation und eine signifikant größere externe Kontrollüberzeugung weisen auf psycho-soziale Problemlagen hin, die systematisch mit den Gründen für die Nicht-Teilnahme an der Befragung am Haupttermin verbunden sind.

Tabelle 5: sozio-demographische Merkmale und Einstellungen von ‚konvertierten' Nicht-Teilnehmern der Schulbefragung (Köln)

% - Anteile	**‚Abwesende'**[a]		**‚Schulverweigerer'**[b]	
	Haupttermin	Nachholtermin	Kontrollgruppe	Schulverweigerer
Stichprobengröße n=	264	27	1045	38
Geschlecht: männlich	53,2	29,6*	55,1	70,3(*)
hoher / sehr hoher Berufsstatus	11,4	20,0	11,6	0,0*
elterl. Bildungsabschluss Abitur oder mehr	9,5	14,8	9,1	7,9
Arbeitslosigkeit	34,1	33,3	29,4	36,0
Prävalenz Schuleschwänzen (letztes Jahr)	36,5	50,0	36,2	67,7***
Positive Schulmotivation[c]	2,2	2,0	2,2	2,0**
Externe Kontrollüberzeugung[c]	0,9	1,3*	1,0	1,3***
Befürwortung von Normverletzungen[c]	1,2	1,1	1,2	1,6**

a Teilstichprobe; Befragte an fünf Kölner Hauptschulen, an denen ein Nachholtermin für am Haupttermin abwesende SchülerInnen angeboten wurde.
b Schulverweigerer: 38 Personen in sozialpädagogischen ‚Schulschwänzer-Projekten'; Kontrollgruppe: alle Befragten an Sonder- und Hauptschulen
c Mittelwerte einer Skala von 0-3

Etwas deutlicher fallen die Unterschiede zwischen der Gruppe der ‚Schulverweigerer', die in den allgemeinbildenden Schulen überhaupt nicht mehr erreicht werden können, und der Kontrollgruppe der Haupt- und Sonderschüler aus.[13] Die Schulverweigerer sind

13 Da 61,5 Prozent der Befragten in den Schulschwänzerprojekten zuvor eine Sonder- oder Hauptschule und weitere 28,8 Prozent eine Gesamtschule besucht haben, und sich in der Kölner Stichprobe keine Gesamtschüler befinden, werden die Sonder- und Hauptschüler als Kontrollgruppe definiert.

überwiegend männlich, das elterliche Berufsprestige ist signifikant niedriger als in der Kontrollgruppe. Erwartungsgemäß berichten erheblich mehr Befragte in der Gruppe der Schulverweigerer über Schulschwänzen, und bei Schulmotivation, externer Kontrollüberzeugung und auch bei der Akzeptanz devianter Normen zeigen sich hochsignifikante Unterschiede zur Kontrollgruppe, die für eine erhebliche Mehrbelastung mit psycho-sozialen Problemen sprechen.

Während also die Befragung der am Haupttermin abwesenden, aber grundsätzlich teilnahmebereiten Schüler keine Hinweise auf sozio-demographische Stichprobenverzerrungen durch Nonresponse bei Schulbefragungen ergab, legt eine Befragung dauerhafter Schulverweigerer ein anderes Ergebnis nahe. Zumindest für diese kleine Gruppe, deren Umfang unbekannt ist, gilt, dass ihre faktische Nicht-Erreichbarkeit bei Schulbefragungen für eine systematische Verzerrung der Stichproben und der Befragungsergebnisse sorgt.

2.3 Angaben zur selbstberichteten Delinquenz

2.3.1 Vergleich der Prävalenzraten

Im folgenden werden die Prävalenzraten der selbstberichteten Delinquenz (SRD) bezogen auf die letzten zwölf Monate aus beiden Befragungen miteinander verglichen. Die verwendeten vierzehn Einzelitems aus der SRD-Skala betreffen ausnahmslos strafbare Verhaltensweisen und werden anhand inhaltlicher Nähe zu Deliktsgruppen zusammengefasst (siehe auch Anhang). Um zu prüfen, inwieweit sozio-demographische Ungleichheiten der Stichproben zu unterschiedlichen Prävalenzraten führen, sind in Tabelle 6 die Prävalenzraten der Gesamtdelinquenz zunächst nach Alter, Schultyp und ethnischer Herkunft der Jugendlichen getrennt dargestellt.

Es zeigt sich, dass für fast alle Subgruppen die Prävalenz in der mündlichen Befragung deutlich niedriger liegt als die Prävalenz in der Schulbefragung. Insgesamt liegt die Prävalenzrate bei der mündlichen Befragung um ca. ein Viertel niedriger als bei der Schulbefragung. Die Unterschiede fallen bei den Schultypen mit niedrigerem Bildungsstauts (Sonder-, Haupt-, Realschule) größer aus als bei den Schultypen mit höherem Bildungsstatus (Gymnasium, Waldorfschule), wo sie nicht signifikant sind. Die besonders großen Differenzen bei den Sonderschülern sind wegen sehr niedriger Fallzahlen allerdings statistisch nicht bedeutsam. Das hat entsprechende Konsequenzen für die Höhe des Zusammenhangs zwischen Schulform und Prävalenzrate in beiden Stichproben: Während die Zwischengruppen-Varianz in der Schulbefragung mit Eta=0.22 (p=0.002) berechnet wird, erreicht sie in der mündlichen Befragung nur Eta=0.05 und ist nicht signifikant (p=0.956); ein ähnlicher Unterschied (Eta = .41 vs. .25) tritt bei den Prävalenzraten der schweren Delinquenz (schwere Eigentums- und Gewaltdelikte)

auf.[14] Je nach Befragungsmethode erhält man demnach unterschiedliche Antworten auf die Frage, ob und wie stark sich die Prävalenzraten der Jugenddelinquenz nach Schulformen unterscheiden.

Tabelle 6: Gesamtprävalenz der Delinquenz (in %) nach soziodemographischen Merkmalen und Erhebungsform

Prävalenz (%)		Schulbefragung	mündliche Befr.	Differenz[a]
Herkunft	nicht-deutsch	80,7	55,7	31,0***
	deutsch	69,2	54,9	20,7***
Alter	15	68,5	56,2	18,0*
	16	77,1	54,1	29,9***
Schultyp	Sonderschule	72,0	33,3	53,8
	Hauptschule	86,7	59,3	31,6***
	Realschule	81,2	55,8	31,3***
	Gesamtschule	66,1	52,2	21,0
	Gymnasium	64,0	54,0	15,6
	Waldorfschule	58,3	60,0	-2,9
Gesamte Stichprobe		72,0	55,1	23,5***

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

a Die Differenz ergibt sich aus folgender Berechnung: (Rate Schulbefr. – Rate mündl. Befr.) / Rate mündl. Befr. * 100.

Bei einem nach Deliktsgruppen differenzierten Vergleich bestätigen sich diese Unterschiede (Tabelle 7). Bei dieser Auswertung werden Befragte mit hohem Bildungsniveau (Gymnasiasten und Waldorfschüler) einerseits und Befragten mit niedrigem Bildungsniveau (Schüler aller übrigen Schulformen) andererseits unterschieden. Die Differenzen zwischen den Prävalenzraten in beiden Befragungen sind in der Subgruppe der Befragten mit niedrigem Bildungsstatus durchweg stärker als in der Subgruppe der Befragten mit hohem Bildungsstatus. Für beide Subgruppen gilt jedoch, das die Differenzen bei eher leichten Deliktsformen (z.B. einfacher Diebstahl, Drogendelikte) relativ geringer und bei eher schweren Deliktsformen (z.B. schwerer Diebstahl) relativ größer ausfallen. Zusätzlich wird hier die Prävalenzrate des Schulschwänzens berichtet, die im Gegensatz zu allen strafbaren Handlungen in der mündlichen Befragung höher als in der Schulbefragung liegt. Dies erklärt sich unseres Erachtens aus der geringeren Chance, Schulschwänzer in

14 Sonderschüler wurden in beiden Stichproben ausgeschlossen.

der Schule anzutreffen; hier scheint eine haushaltsbasierte Stichprobenziehung von Vorteil zu sein.[15]

Tabelle 7: Prävalenzraten nach nach Deliktsgruppen, Schulform und Erhebungsform

Schulform	Deliktsgruppen	Schulbefragung	mündl. Befragung	Differenz[a]
Förder-, Haupt-, Real-, Gesamtschule	Sachbeschädigung	39,1	24,9	36,3**
	Einfacher Diebstahl	60,4	35,9	40,6***
	Drogendelikte	33,8	19,7	41,7**
	Schwerer Diebstahl	24,3	13,0	46,5**
	Gewaltdelikte	39,0	19,3	50,5***
	Alle Delikte	77,9	56,1	28,0***
	Schulschwänzen	35,4	41,6	-17,5
Gymnasium, Waldorfschule	Sachbeschädigung	31,5	23,6	25,1
	Einfacher Diebstahl	33,7	24,6	27,0
	Drogendelikte	37,7	27,8	26,3
	Schwerer Diebstahl	5,7	3,0	47,4
	Gewaltdelikte	13,2	10,9	17,4
	Alle Delikte	63,5	54,4	14,3
	Schulschwänzen	30,1	33,3	-10,6

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$ (T-Test)

a Die Differenz ergibt sich aus folgender Berechnung: (Rate Schulbefr. – Rate mündl. Befr.) / Rate mündl. Befr. * 100.

Eine Besonderheit verbirgt sich hinter den Unterschieden zwischen Befragten deutscher und nicht-deutscher Herkunft. Während die Differenz zwischen den Prävalenzraten der Gesamtdelinquenz der beiden Befragungen für die Gruppe der nicht-deutschen Befragten höher ausfällt (31,0 Prozent vs. 20,7 Prozent, siehe Tabelle 6), besteht bei den Prävalenzraten der Deliktsgruppen schwerer Diebstahl, Gewalt- und Drogendelikte ein umgekehrtes Verhältnis (Tabelle 8). Dies hat zur Folge, dass die Prävalenzraten der nicht-deutschen Befragten in der Deliktsgruppe schwerer Diebstahl nur bei der mündlichen Befragung – also umgekehrt wie beim Merkmal besuchte Schulform – signifikant höher sind als die der deutschen Befragten. In einem weiteren inhaltlich bedeutsamen Aspekt weichen demnach die Ergebnisse von Schulbefragung und mündlicher Befragung voneinander ab. Dieses Resultat widerspricht den Erkenntnissen bisheriger empirischer Studien zur Validität selbstberichteter Delinquenz,

15 Der Zusammenhang zwischen Schulschwänzen und strafbarer Delinquenz ist mit rho=.36 in der Stichprobe der Schulbefragung und rho=.33 in der Stichprobe der mündlichen Befragung nicht sehr eng.

wonach Delikte häufiger von ethnischen Minoritäten als von Einheimischen verschwiegen werden (Junger 1989; Hindelang u.a. 1981; Huizinga/Elliott 1986).

Tabelle 8: Prävalenzraten nach Deliktsgruppen, ethnischer Herkunft und Erhebungsform

	Deliktsgruppen	Schul-befragung	Mündliche Befragung	Differenz[a]
Deutsche[b]	Sachbeschädigung	36,1	25,5	29,4**
	Schwerer Diebstahl	15,3	6,6	56,9***
	Einfacher Diebstahl	45,1	28,3	37,3***
	Drogendelikte	37,7	23,8	36,9***
	Gewaltdelikte	24,9	12,9	48,2***
Nicht-Deutsche[b]	Sachbeschädigung	36,4	18,0	50,5**
	Schwerer Diebstahl	21,6	15,2	29,6
	Einfacher Diebstahl	63,2	38,5	39,1***
	Drogendelikte	28,2	23,4	17,0
	Gewaltdelikte	37,9	26,6	29,8

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

a Die Differenz ergibt sich aus folgender Berechnung: (Rate Schulbefr. – Rate mündl. Befr.) / Rate mündl. Befr. * 100.

b Deutsche: Herkunftsland von Vater oder Mutter Deutschland; Nicht-Deutsche: Herkunftsland von Vater und Mutter nicht Deutschland

2.3.2 Vergleich der Zusammenhänge mit anderen Variablen

Wie wirken sich unterschiedliche Erhebungsmethoden auf die für die Delinquenzsoziologie zentralen Fragen nach den Einflussfaktoren auf jugendliche Delinquenz auf? Es wurde bei den Gruppenvergleichen bereits deutlich, dass Zusammenhänge von selbstberichteter Delinquenz mit den strukturellen Variablen Ethnie und besuchter Schulform je nach Befragungsart unterschiedlich ausfallen. Gleiches gilt auch für den Vergleich von besuchter Schulform und elterlichem Berufsprestige: Während der Zusammenhang zwischen schwerer Delinquenz und besuchter Schulform in der Schulbefragung stärker ist als in der mündlichen Befragung (tau-b =-.37 vs. -.18), ist der Zusammenhang mit dem elterlichen Berufsprestige in der Schulbefragung schwächer als in der mündlichen Befragung (tau-b= -.10 vs. -.21). Ein Vergleich bivariater Korrelationskoeffizienten der Delinquenz mit einer Reihe weiterer relevanter Variablen ergibt teils sehr geringe, teils deutliche Unterschiede; in keinem Falle jedoch kehrt sich die Richtung eines Zusammenhanges um (Tabelle 9).

Tabelle 9: bivariate Korrelationen zwischen Delinquenz (Inzidenzraten) und familien- bzw. schul-/peerbezogenen Variablen

bivariater Korrelations-koeffizient (Spearman)	Gewaltdelikte		schwere Eigentumsdelikte		Drogendelikte	
	Schulb.	mündl.	Schulb.	mündl.	Schulb.	mündl.
Familienbezogene Variablen						
Familienform (0=vollständig, 1=unvollst.)	,01	-,02	,10*	,14*	,08	,09
Ethnische Herkunft (0=deutsch., 1=nicht-dt..)	,14**	,19***	,07	,13*	-,08	-,04
elterl. Berufsprestige (1-4)	-,13*	-,18**	-,04	-,21**	,17**	,06
elterl. Bildungsstatus (1-4)	-,15**	-,15**	-,07	-,13*	.18***	,13*
Streit mit Eltern[a]	,17*	,17**	,07	,21***	,16*	,18**
Schul-/peerbezogene Variablen						
Schultyp[b]	-,35***	-,14*	-,29***	-,21***	,08	.07
Positive Schulmotivation[a]	-,15**	,02	-,14**	-,04	-,31***	-,14*
delinquente peers[a]	,46***	,25***	,43***	,32***	,53***	,39***
Mitglied in delinquenter Clique (0=nein, 1=ja)	,50***	,29***	,34***	,32***	,21***	,06
Freizeitorientierung 'Unterhaltung/Action'[a]	,48***	,22***	,41***	,14*	,47***	,25***

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$; Delikte: Inzidenz letztes Jahr; [a] Intervallskalen

a intervallskalierte Skalen (Wertebereich 0-2 bzw. 0-3)

b 1=Sonder-, Hauptschule, 2=Real-, Gesamtschule, 3=Gymnasium, Waldorfschule

grau hinterlegte Felder: jeweils höhere Koeffizienten

Insgesamt deutet sich ein Muster an, nach dem Zusammenhänge von Delinquenz mit familienbezogenen Variablen in der mündlichen Befragung stärker, und Zusammenhänge von Delinquenz mit schul- und peerbezogenen Variablen in der Schulbefragung stärker in Erscheinung treten. Dies gilt sowohl für strukturelle Variablen als auch noch stärker für Einstellungsvariablen wie etwa Schulmotivation und Freizeitorientierungen. (vgl. dazu ausführlich Naplava/Oberwittler 2002).

2.3.3 Zusammenfassung und Diskussion

Kurz zusammengefasst hat sich gezeigt, dass die Prävalenzraten selbstberichteter Delinquenz bei der haushaltsbasierten mündlichen Befragung niedriger liegen als bei der schulbasierten schriftlichen Befragung; dies gilt jedoch weniger für die Subgruppe der Befragten mit hohem Bildungsstatus als für die Subgruppe der Befragten mit niedrigem Bil-

dungsstatus. Gleichzeitig besteht in beiden Subgruppen die Tendenz, dass die Differenz der Prävalenzraten bei schweren Deliktsarten größer ist als bei leichten Deliktsarten. Außerdem ergab sich der unerwartete Befund, dass diese Differenzen in der Gruppe der nicht-deutschen Befragten teilweise deutlich geringer sind als in der Gruppe der deutschen Befragten.

Ob diese Unterschiede eher auf Selektionseffekte der Erhebungsverfahren oder auf Antworteffekte der Befragungssituationen zurückzuführen sind, lässt sich anhand des hier durchgeführten Vergleiches nicht eindeutig entscheiden. Uns erscheint die Annahme plausibel, dass die Selektionseffekte für den größeren Teil der Differenzen verantwortlich sind, d.h. dass bei einer haushaltsbasierten Befragung weniger delinquente Jugendliche erreicht werden als bei der schulbasierten Befragung. Dafür spricht über die erheblich geringere Ausschöpfungsrate der haushaltsbasierten Befragung hinaus, dass die Zusammensetzung der haushaltsbasierten im Vergleich zur schulbasierten Stichprobe nicht nur sozialstrukturell, sondern auch hinsichtlich der besuchten Schulformen zu den oberen Statusgruppen hin verzerrt ist. Jugendliche, die Sonder- und Hauptschulen besuchen, sowie Jugendliche aus sozial benachteiligten Stadtteilen wurden in der haushaltsbasierten Stichprobe relativ schlechter ausgeschöpft als in der schulbasierten Stichprobe. Aus der bisherigen Forschung ist bekannt, dass sich in diesen Gruppen mehr delinquente Jugendliche befinden als in anderen Gruppen. Kontrolltheoretisch lässt sich begründen, warum Jugendliche mit devianten Orientierungen seltener an Befragungen teilnehmen als andere. Wenn man annimmt, dass die Delinquenzneigung ein Selektionskriterium für die Teilnahme an der mündlichen Befragung darstellt, dann ist auch der Befund plausibel, dass die Differenz der Prävalenzraten bei schweren Deliktsformen und der Intensivtäter größer ist als bei leichten Deliktsformen. Denn leichte Delikte werden von sehr vielen Jugendlichen begangen und gehen meist nicht mit einer starken Delinquenzneigung einher; bei schweren Delikten ist diese ausgeprägter, so dass angenommen werden kann, dass insbesondere Jugendliche, die schwere Delikte begehen, nicht an der mündlichen Befragung teilnehmen.

Für die ‚Selektions-These' spricht auch der Befund, dass die Differenzen der Prävalenzraten bei den deutschen Jugendlichen wesentlich stärker sind als bei den nicht-deutschen, denn die Selektivität der haushaltsbasierten Stichprobenziehung betraf ebenfalls vorrangig die deutschen Jugendlichen. Auch wenn nicht klar ist, warum nicht-deutsche Jugendliche keine Selektivität nach Stichprobenverfahren zeigen, so korrespondiert dies doch mit gleichfalls geringen Unterschieden in den Befragungsergebnissen.

Schließlich weist auch der gegenläufige Effekt beim Schulschwänzen auf die Existenz von Selektionseffekten hin; denn hier lässt sich die geringere Prävalenzrate bei der Schul-

befragung mit der geringeren Wahrscheinlichkeit erklären, Schulschwänzer in der Schule anzutreffen.

Anders sieht es unseres Erachtens bei der Frage unterschiedlicher Zusammenhangsmuster der Delinquenz aus. Gegen die Selektions-These sprechen die gegenläufigen Effekte, die in beiden Befragungen bei den Zusammenhängen der Delinquenz mit unterschiedlichen Schichtindikatoren feststellbar sind. Obwohl das haushaltsbasierte Erhebungsverfahren zu einer geringeren Ausschöpfung in den unteren Statusgruppen geführt hat und angenommen werden kann, dass damit auch Jugendliche mit delinquenten Neigungen schlechter ausgeschöpft wurden als bei der schulbasierten Erhebung, ist der Zusammenhang zwischen selbstberichteter Delinquenz und elterlichem Berufsprestige nicht schwächer, sondern sogar stärker als in der Schulbefragung. Dass dieses Muster darüber hinaus auch Einstellungsvariablen gilt, legt die – zugegeben spekulative – Vermutung nahe, dass hierfür ein Effekt des Erhebungskontextes auf die (unbewusste) Bewertung der ‚Salienz' von Einflussfeldern auf das eigene delinquente Verhalten beteiligt ist. Dies würde bedeuten, dass Jugendliche im Erhebungskontext der Schule den schulischen Faktoren sowie den Beziehungen zu den Gleichaltrigen eine höhere Wertigkeit beimessen als den familiären Faktoren, während diese familiären Faktoren für die Befragten im Erhebungskontext des elterlichen Haushalts eine größere Wertigkeit besitzen. Möglicherweise übt die Anwesenheit der anderen Jugendlichen während der Schulbefragung einen Druck der sozialen Erwünschtheit in der Art aus, das Antwortverhalten zu Fragen, die die Gleichaltrigen betreffen, an den jeweils gültigen Gruppennormen (ob deviant oder nicht-deviant) auszurichten. In Schulklassen, in denen die Gruppennormen überwiegend nicht-deviant sind (zum Beispiel an Gymnasien), würden die Befragten ihre delinquenten Einstellungen und Handlungen eher untertreiben, und in Schulklassen, in denen die Gruppennormen überwiegend delinquent sind (zum Beispiel an Hauptschulen), würden die Befragten ihre Delinquenz eher noch übertreiben. Ein solcher Kontexteffekt der Befragungssituation könnte zu einer größeren Homogenität innerhalb der Klassen und Schultypen und zu stärkeren Korrelationen der peerbezogenen Variablen mit der abhängigen Variable führen.

3. Schlussfolgerungen

Das wesentliche Ergebnis unseres Methodenvergleichs ist, dass Befragungsergebnisse zur selbstberichteten Delinquenz spürbar vom Erhebungsverfahren beeinflusst werden und daher entsprechend vorsichtig interpretiert werden sollten – dies um so mehr, als davon nicht nur die Prävalenzen, sondern auch die Zusammenhangsmuster betroffen sind. Insgesamt sprechen die Ergebnisse des Methodenvergleichs dafür, dass Schulbefragungen wegen ihrer sehr hohen Ausschöpfungsraten und der damit verbundenen geringen Selek-

tivität bei Studien zur selbstberichteten Delinquenz gegenüber haushaltsbasierten Erhebungsverfahren validere Ergebnisse produzieren. Daher erscheinen uns Schulbefragungen als ein empfehlenswertes Erhebungsverfahren, wenn die Zielpopulation Jugendliche im schulpflichtigen Alter (de facto bis sechzehn Jahre) sind.

Da wir jedoch mögliche Hinweise auf Kontexteffekte der Befragungssituation gefunden haben, scheinen auch die Ergebnisse von Schulbefragungen nicht frei von spezifischen Verzerrungen zu sein. Weil es nur wenige Untersuchungen zu den Auswirkungen von Schulbefragungen im Gruppenkontext auf das Antwortverhalten gibt, fehlt es an der notwendigen empirischen Basis zur Überprüfung dieses Verdachts.

Notwendig erscheinen uns in dieser Situation weitere Methodenstudien, bei denen der Befragungskontext bei Schulbefragungen experimentell variiert wird, so dass der vermutete Effekt der gegenseitigen Beeinflussung der Befragten kontrolliert werden kann. Hierbei könnten auch die bislang wenig erforschten Auswirkungen computerbasierter Befragungstechniken (CASI) bei Gruppenbefragungen auf das Antwortverhalten untersucht werden. Bisherige Studien haben gezeigt, das der Einsatz von CASI gegenüber PAPI zu höheren Prävalenzraten bei Fragen nach ‚heiklen' Verhaltensweisen führt (Beebe et al. 1998; Hallfors et al. 2000; De Leeuw et al. 1997; Turner et al. 1998). Denkbar wäre unseres Erachtens auch, das der Technikeinsatz zu einem erhöhten Gefühl der Vertraulichkeit führt – unter anderem durch die Verwendung von Kopfhörern bei Audio-CASI und durch die erschwerte Einsehbarkeit der Antworten auf dem Computerdisplay – , und dadurch mögliche Antworteffekte des Gruppenkontextes reduziert werden könnten.

Korrespondenzadresse

Dr. Dietrich Oberwittler
Max-Planck-Institut für ausländisches und internationales Strafrecht Freiburg i.Br.
Günterstalstraße 73
D-79100 Freiburg
Tel. 0761 / 7081-250
d.oberwittler@iuscrim.mpg.de

Literaturverzeichnis

Albrecht, G./Howe, C.-W./Wolterhoff-Neetix, J., 1988: Neue Ergebnisse zum Dunkelfeld der Jugenddelinquenz: Selbstberichtete Delinquenz von Jugendlichen in zwei westdeutschen Großstädten. S. 661-696 in: Kaiser, G. (Hg.), Kriminologische Forschung in den 80er Jahren. Freiburg: Max-Planck-Institut für ausländisches und internationales Strafrecht.

Beebe, T.J./Harrison, P.A./McRae, J.R./Anderson, R.E./Fulkerson, J.A., 1998: An evaluation of computer-assisted self-interviews in a school setting. Public Opinion Quarterly 62, 623-632.

Currie, C./Hurrelmann, K./Settertollubte, W./Smith, R./Todd, J., 2000: Health and Health Behaviour Among Young People: A WHO Cross-National Study (HBSC). Kopenhagen: WHO.

Deutsche Forschungsgemeinschaft/Kaase, M. (Hg.), 1999: Qualitätskriterien der Umfrageforschung. Denkschrift. Berlin: Akademie-Verlag.

Deutsche Shell (Hg.) 2000. Jugend 2000: 13. Shell Jugendstudie. Opladen: Leske + Budrich.

Deutsche Shell (Hg.) 2002. Jugend 2002: Zwischen pragmatischem Idealismus und robustem Materialismus. Frankfurt a.M.: Fischer.

Diekmann, A., 1998: Empirische Sozialforschung: Grundlagen, Methoden, Anwendungen. Reinbeck bei Hamburg: Rowohlt.

Dillman, D. A., 1983: Mail and other Self-Administered Questionnaires. S. 359-376 in: Rossi, P. H./Wright, J. D./Anderson, A. B., Handbook of Survey Research. New York: Academic Press.

Friedrichs, J., 1990: Methoden empirischer Sozialforschung, 14. A.. Opladen: Westdeutscher Verlag.

Fuchs, M./Lamnek, S./Luedtke, J., 2001: Tatort Schule: Gewalt an Schulen 1994-1999. Opladen: Leske + Budrich.

Hallfors, D./Khatapoush, S./Kadushin, C./Watson, K./Saxe, L., 2000: A comparison of paper vs. computer-assisted self interview for school alcohol, tobacco, and other drug surveys. Evaluation and Program Planning 23: 149-155.

Heitmeyer, W./Collmann, B./Conrads, J./Matuschek, I./Kraul, D./Kühnel, W./Möller, R./Ulbrich-Hermann, M., 1996: Gewalt: Schattenseiten der Individualisierung bei Jugendlichen aus unterschiedlichen Milieus. Weinheim/München: Juventa.

Hindelang, M./Hirschi, T./Weis, J., 1981: Measuring Delinquency. Beverly Hills: Sage.

Hirschi, T., 1969: Causes of Delinquency. Berkeley: University of California Press.

Holtappels, H. G./Heitmeyer, W./Melzer, W./Tillmann,K.-J. (Hg.), 1997: Forschung über Gewalt an Schulen: Erscheinungsformen und Ursachen, Konzepte und Prävention. Weinheim/München: Juventa.

Huizinga, D./Elliott, D,. 1986: Reassessing the reliability and validity of self-report delinquency measures. Journal of Quantitative Criminology 2: 293-327.

Junger, M., 1989: Discrepancies Between Police and Self-Report Data for Dutch Racial Minorities. British Journal of Criminology, 29(3): 273-284.

Kerkvliet, J., 1994: Cheating by Economics Students: A Comparison of Survey Results. Journal of Economic Education, 25(2): 121-133.

Klein, S./ Porst, R., 2000: Mail Surveys. Ein Literaturbericht (ZUMA Technischer Bericht 10/2000). Mannheim: Zentrum für Umfragen, Methoden und Analysen.

Kreuzer, A./Görgen, T./Römer-Klees, B./Schneider, H., 1992: Auswirkungen unterschiedlicher methodischer Vorgehensweisen auf die Ergebnisse selbstberichteter Deliquenz. Monatsschrift für Kriminologie und Strafrechtsreform 75(2/3): 91-104.

De Leeuw, E./Hox, J./Kef, S./van Hattum, M., 1997: Overcoming the Problems of Special Interviews on Sensitive Topics: Computer Assisted Self-Interviewing Tailored for Young Children and Adolescents. S. 1-14 in: Sawtooth Software Inc. (Hg.), Sawtooth Software Conference Proceedings. Sequim, WA: Sawtooth Software.

Mansel, J., 2001: Angst vor Gewalt: Eine Untersuchung zu jugendlichen Opfern und Tätern. Weinheim; München: Juventa.

Mansel, J./Hurrelmann, K., 1998: Aggressives und delinquentes Verhalten Jugendlicher im Zeitvergleich. Befunde der 'Dunkelfeldforschung' aus den Jahren 1988, 1990 und 1996. Kölner Zeitschrift für Soziologie und Sozialpsychologie, 50(1): 78-109.

Merkens, H., (Hg.) 1999: Schuljugendliche in beiden Teilen Berlins seit der Wende: Reaktionen auf den sozialen Wandel. Baltmannsweiler: Schneider-Verl. Hohengehren.

Naplava, T./Oberwittler, D., 2002: Methodeneffekte bei der Messung selbstberichteter Delinquenz von männlichen Jugendlichen – Ein Vergleich zwischen schriftlicher Befragung in der Schule und mündlicher Befragung im Haushalt. Monatsschrift für Kriminologie und Strafrechtsreform, 85(6), im Druck.

Oberwittler, D./Blank, T./Köllisch, T./Naplava, T., 2001: Soziale Lebenslagen und Delinquenz von Jugendlichen. Ergebnisse der MPI-Schulbefragung 1999 in Freiburg und Köln. Freiburg: edition iuscrim.

Oberwittler, D./Köllisch, T./Würger, M., 2002: Selbstberichtete Delinquenz bei Jugendlichen. In: Glöckner-Rist, A. (Hg.), ZUMA-Informationssystem. Elektronisches Handbuch sozialwissenschaftlicher Erhebungsinstrumente Version 6.0. Mannheim: Zentrum für Umfragen, Methoden und Analysen.

Reuband, K.-H., 1985: Methodische Probleme bei der Erfassung altersspezifischer Verhaltensweisen: die Zusammensetzung von Interviewerstäben und ihr Einfluss auf das Antwortverhalten Jugendlicher. ZA-Information, 17: 34-50.

Reuband, K-H./Blasius, J., 1996: Face-to-face-, telefonische und postalische Befragungen: Ausschöpfungsquoten und Antwortmuster in einer Großstadt-Studie. Kölner Zeitschrift für Soziologie und Sozialpsychologie, 46(2): 296-318.

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

Schnell, R./ Hill, P./Esser, E., 1999: Methoden der empirischen Sozialforschung, 6.A. München: Oldenbourg.

Schumann, K. F./Berlitz, C./Guth, H.-W./Kaulitzki, R., 1987: Jugendkriminalität und die Grenzen der Generalprävention. Neuwied: Luchterhand.

Sturzbecher, D., (Hg.), 2001: Jugend in Ostdeutschland: Lebenssituation und Delinquenz. Opladen: Leske + Budrich.

Tillmann, H.-J./Holler-Nowotzki, B./Holtappels, H. G./Meier, U./Popp, U., 1999: Schülergewalt als Schulproblem: verursachende Bedingungen, Erscheinungsformen und pädagogische Handlungsperspektiven. Weinheim/München: Juventa.

Turner, C.F./Ku, L./Rogers, S.M./Lindberg, L.D./Pleck, J.H./Sonenstein, F.L., 1998: Adolescent Sexual Behavior, Drug Use, and Violence: Increased Reporting with Computer Survey Technology. Science 280: 867-873.

Villmow, B./ Stephan, E., 1983: Jugendkriminalität in einer Gemeinde: Eine Analyse erfragter Delinquenz und Viktimisierung sowie amtlicher Registrierung. Freiburg: Max-Planck-Institut.

Wegener, B., 1988: Kritik des Prestiges. Opladen: Westdeutscher Verlag.

Wetzels, P./Enzmann, D./Pfeiffer, C., 2001: Jugend und Gewalt. Eine repräsentative Dunkelfeldanalyse in München und acht anderen deutschen Städten. Baden-Baden: Nomos.

Zinnecker, J./Behnken, I./Maschke, S./Stecher, L. 2002: Null zoff & voll busy. Die erste Jugendgeneration des neuen Jahrhunderts. Ein Selbstbild. Opladen: Lese + Budrich.

Anhang

Skala: ‚Selbstberichtete Delinquenz Jugendlicher' (Oberwittler et al. 2002)

Die meisten Menschen tun in Ihrem Leben manchmal Dinge, die verboten sind, z.B. ohne Fahrkarte im Bus fahren oder etwas stehlen. Wir möchten gerne von Dir wissen, ob Du auch schon mal etwas Verbotenes getan hast.

Ich habe in den letzten zwölf Monaten (alleine oder zusammen mit anderen) …

(Sachbeschädigung)
... mit einer Spraydose irgendwo Sprüche oder Bilder aufgesprüht (Graffiti).
... etwas absichtlich in der Schule, in Parks, Telefonzellen, der U-Bahn usw. beschädigt oder zerstört.
... Autos, Motorräder oder Motorroller usw. absichtlich beschädigt.
(einfacher Diebstahl)
... ein Fahrrad oder ein Teil eines Fahrrades gestohlen (Sattel, Rad)
... in einem Geschäft etwas gestohlen.
... jemandem eine Sache oder Geld gestohlen.
(schwerer Diebstahl)
... ein Auto aufgebrochen.
... ein Auto, Motorrad, Motorroller usw. gestohlen.
... irgendwo eingebrochen, um etwas zu stehlen (in ein Haus, Keller, Laden usw.).
(Drogendelikte)
... Drogen genommen (Haschisch, Ecstasy usw.).
... Drogen an andere verkauft(Haschisch, Ecstasy usw.).
(Gewaltdelikte)
... jemanden so geschlagen oder verprügelt, dass er/sie verletzt war oder blutete.
... jemanden bedroht oder erpresst, um ihm/ihr wirklich Angst zu machen, oder um Geld oder eine bestimmte Sache zu bekommen.
... jemandem mit Gewalt etwas weggenommen (durch Festhalten, Schlagen usw.)

Generierung eines Proxys zum Job-Anforderungsniveau aus den Informationen zu ausgeübtem Beruf und beruflicher Stellung.

Ein neues Tool für die deutsche Überqualifikations-Forschung

Matthias Pollmann-Schult & Felix Büchel

Die Fehlqualifikations-Forschung gewinnt als eigenständige und innovative Teildisziplin der Arbeitsmarktforschung schnell an Bedeutung. Als Fehlqualifikation wird eine fehlende Übereinstimmung zwischen erworbener beruflicher Qualifikation und den Anforderungen der ausgeübten beruflichen Tätigkeit bezeichnet. Von besonderem Interesse dabei ist der vertikale Mismatch in Form einer Überqualifikation (für einen Literaturüberblick vgl. Büchel 2001).

Ein wesentliches Hemmnis der weiteren Entwicklung der deutschen Überqualifikations-Forschung liegt in einer spezifischen Datenrestriktion: Häufig enthalten die verfügbaren Mikrodaten keine hinreichend genauen Angaben zum Job-Anforderungsniveau der Erwerbstätigen. In dieser Situation behelfen sich viele Forscher dadurch, dass sie die Information zur beruflichen Stellung als Proxy für das Job-Anforderungsniveau verwenden. Dieses Verfahren hat jedoch schwer wiegende Nachteile, die die Validität der Messung von Überqualifikation in Frage stellt.

Im vorliegenden Beitrag wird ein leistungsfähigeres Verfahren zur Generierung eines Proxys des Job-Anforderungsniveaus vorgestellt. Dieses basiert auf den Informationen zum ausgeübten Beruf und zur beruflichen Stellung. Beide Variablen sind in den meisten Datensätzen verfügbar. Mit dem neuen Verfahren werden somit auch Datensätze, die das Job-Anforderungsniveau der Erwerbstätigen nicht explizit erfragen, für die Überqualifikationsforschung nutzbar.

The analysis of skill mismatch is gaining rapidly in importance as an original and innovative area in labour market research. The term 'skill mismatch' describes a situation in which the occupational qualifications of employees do not match those required to perform their current job. Of special interest here is vertical skill mismatch in the form of over-qualification.

Specific data restrictions constitute a major impediment to German research on over-qualification in the labour market. The micro data sets available often do not provide sufficiently detailed information on job demand levels, that is, the level of skills required. Many researchers thus turn to information available on people's occupational position and use this as a proxy for the level of skills required. However, this strategy has a number of serious drawbacks that call into question the validity of the measurement of over-qualification. The article describes a more powerful procedure to generate a proxy of the level of required job skills. The input for this procedure is derived from information on the occupational position and the occupation of people in employment, both variables that are available in most data sets. This new procedure makes data sets that do not contain explicit information on the required level of skills better suited for use in job over-qualification research.

1. Einleitung

Alle Messkonzepte der vertikalen Fehlqualifikations-Forschung[1] basieren auf einem Abgleich zwischen formaler beruflicher Qualifikation der Erwerbstätigen und dem Anforderungsniveau des Arbeitsplatzes. Dieser Abgleich indiziert, ob die Befragungsperson ausbildungsadäquat oder aber unter- bzw. überwertig beschäftigt ist. Während Angaben zur beruflichen Qualifikation in allen gängigen Datensätzen verfügbar sind, wird das Anforderungsniveau des Arbeitsplatzes nur selten erhoben. Dieses Fehlen der Information zum Job-Anforderungsniveau in zahlreichen Datensätzen hemmt die deutsche Fehlqualifikations-Forschung erheblich.

Die Fehlqualifikations-Forschung verwendet kein einheitliches Messkonzept des Arbeitsanforderungsniveaus[2]; im Wesentlichen wird zwischen dem sog. „subjektiven“ und dem

---

1 Von einer vertikalen Fehlqualifikation ist eine horizontale abzugrenzen. Horizontale Fehlqualifikation liegt vor, wenn die Qualifikation der Erwerbstätigen und die Arbeitsanforderungen des Arbeitsplatzes zwar hinsichtlich des Qualifikationsniveaus, nicht aber des Qualifikationstyps übereinstimmen. Diese Form der Fehlqualifikation wird im folgenden nicht weiter verfolgt.

2 Diese fehlende Standardisierung behindert die Fehlqualifikation noch stärker als die typologisch verwandte Arbeitslosigkeitsforschung. Unterschiedliche Messkonzepte führen zu stark variierenden Anteilen an fehlqualifizierten Erwerbstätigen; zudem birgt die relativ ungenaue Messung das Risiko,

„objektiven“ Ansatz zur Messung des Job-Anforderungsniveaus unterschieden. Beim subjektiven Ansatz wird das Job-Anforderungsniveau direkt erfragt. Als Antwort ergibt sich beispielsweise: „Für die Ausübung meiner Berufstätigkeit ist ein abgeschlossenes Hochschulstudium erforderlich“. Dieses Messkonzept gilt als robust. Messfehler können allerdings durch eine Fehleinschätzung der tatsächlichen Anforderungen des eigenen Arbeitsplatzes auftreten.

Der objektive Ansatz basiert auf den von Arbeitsmarktforschern anhand der berichteten Berufstätigkeit vorgenommenen Beurteilungen des erforderlichen Qualifikationsniveaus. Empirische Mismatch-Studien für die Vereinigten Staaten können hierzu auf die Arbeitsplatzbeschreibungen eines „Dictionary of Occupational Titles“ (DOT) rekurrieren, wo jedem Beruf ein Anforderungsniveau zugewiesen wird. So wird beispielsweise aus der Information: „Befragungsperson ist als Arzt tätig“ die Antwort: „Für die von der Befragungsperson ausgeübte Berufstätigkeit ist ein abgeschlossenes Hochschulstudium erforderlich“ generiert. Ein wesentlicher Nachteil dieses Ansatzes besteht allerdings darin, dass Veränderungen der Arbeitsplatzanforderungen innerhalb einzelner Berufe nur mit großer zeitlicher Verzögerung in den Arbeitsplatzbeschreibungen berücksichtigt werden. Ein noch größeres Problem resultiert aus der in einigen Berufsgruppen sehr großen intraberuflichen Bandbreite des Job-Anforderungsniveaus. Daher wird häufig argumentiert, dass die Erwerbstätigen selbst noch am ehesten darüber Bescheid wüssten, welche Qualifikation ihre Berufstätigkeit tatsächlich erfordert, auch wenn in Einzelfällen eine Fehleinschätzung berichtet wird. Dieser Argumentation folgend herrscht in der Disziplin mittlerweile darüber Einvernehmen, dass der subjektive Messansatz dem objektiven überlegen ist (Hartog 2000). Auf einen objektiven Ansatz wird folglich vorzugsweise dann rekurriert, wenn in der auszuwertenden Datenbasis das Job-Anforderungsniveau nicht direkt erfragt wurde.

Während das objektive Messkonzept in US-amerikanischen Studien relativ häufig angewandt wird, gilt dies kaum für Deutschland. Der Grund liegt darin, dass für Deutschland kein genormter „Dictionary of Occupational Titles“ mit dazugehörenden Anforderungsniveaus verfügbar ist. Dies behindert die Fehlqualifikations-Forschung in Deutschland erheblich, da in den meisten verfügbaren Datenbasen das Job-Anforderungsniveau nicht

dass der Anteil an Fehlqualifizierten systematisch überschätzt wird. Die Forscher-Gemeinde ist sich dieses methodologischen Defizits bewusst und arbeitet entsprechend intensiv an einer Verbesserung der Messtechnik (für einen Überblick über die aktuelle Diskussionslage vgl. Büchel 2001).

direkt erfragt wird.[3] Häufig wird bei dieser prekären Datenlage ein Proxy zum Job-Anforderungsniveau aus der Angabe zur beruflichen Stellung generiert (vgl. zum Beispiel von Henninges 1996; Konietzka 1999; Plicht/Schober/Schreyer 1994). Dabei ist evident, dass diese Vorgehensweise mit erheblichen Validitätsproblemen behaftet ist: Das tatsächliche Job-Anforderungsniveau kann innerhalb der – meist nur hoch aggregiert erhobenen – beruflichen Stellungen erheblich variieren. Im Ergebnis führt dieses Vorgehen häufig zu einer erheblichen Überschätzung der ausbildungsinadäquaten Erwerbstätigkeit in Deutschland.[4] Die Problematik einer Operationalisierung des Job-Anforderungsniveaus über die berufliche Stellung wird damit evident.

Ziel des vorliegenden Beitrags ist es, der deutschen Skill Mismatch-Forschung ein Messkonzept zur Verfügung zu stellen, mit Hilfe dessen die Ausbildungsadäquanz der Beschäftigung anhand der Arbeitsplatzmerkmale „ausgeübter Beruf" und „berufliche Stellung" annähernd gleich valide festgestellt werden kann wie mit Hilfe eines subjektiven Messkonzepts. Dieses Verfahren könnte damit die deutsche Skill-Mismatch-Forschung von der Fixierung auf die wenigen Datensätzen, in denen das Job-Anforderungsniveau direkt erfragt wurde, befreien und so vielfältige neue Forschungsperspektiven eröffnen.

2. Beschreibung des Generierungs-Konzepts

Zur Entwicklung eines Zuordnungsschemas, welches das Job-Anforderungsniveau anhand objektiver Arbeitsplatzcharakteristika indiziert, wird ein Datensatz ausgewählt, der bei hoher Fallzahl sowohl die beiden benötigten Arbeitsplatzmerkmale „ausgeübter Beruf" und „berufliche Stellung" als auch eine „subjektive", das heißt direkt erfragte Angabe zum Job-Anforderungsniveau enthält. Diese Bedingungen werden vom Sozio-oekonomische Panel SOEP erfüllt.[5] Zur Entwicklung des Zuordnungsschemas wird jeder Kombi-

---

3 Ausnahmen bilden das SOEP, einige der BIBB/IAB-Erhebungen, sowie die 1964- und 1971er Kohorten der Westdeutschen Lebensverlaufsstudie des Max-Planck-Instituts für Bildungsforschung, Berlin.

4 So lässt beispielsweise der Befund von Konietzka (1999: 387), wonach 92 Prozent der Fachhochschulabsolventinnen einen Berufseinstieg nur über eine unterwertige Beschäftigung realisieren konnten, zumindest stutzen. Bei einer näheren Betrachtung der Ergebnisse zeigt sich, dass nahezu alle als überqualifiziert kategorisierten Berufsanfängerinnen in die berufliche Stellung „mittlere Angestellte" fallen (85 Prozent). Diese Kategorie ist bekannt als diejenige mit der größten Bandbreite der beruflichen Anforderungen. Es spricht einiges für die Vermutung, dass diese Kategorie auch Tätigkeiten beinhaltet, die durchaus adäquat für Berufsanfängerinnen mit einem Fachhochschulabschluss sind.

5 Für Informationen zu dieser Datenbasis vgl. SOEP Group (2001).

nation von ausgeübtem Beruf – kodiert anhand des ISCO68 [6] – und beruflicher Stellung das „übliche" Job-Anforderungsniveau zugeordnet; „üblich" operationalisieren wir dabei über „am häufigsten berichtet"[7]. Bei dieser Vorgehensweise kann jedoch die Berücksichtigung von Kombinationen mit nur geringen Fallzahlen Probleme bereiten. Es besteht die Gefahr, dass relativ wenige Fälle zu einer Verzerrung des zu generierenden Zuordnungsschemas führen können. Um die Robustheit unseres Zuordnungsschemas zu testen, berechnen wir drei Varianten dieses Schemas. Variante I berücksichtigt nur Kombinationen von ausgeübtem Beruf und beruflicher Stellung mit einer Fallzahl von wenigstens 5 Fällen; bei den Varianten II und III liegt der entsprechende Grenzwert bei 10 bzw. 20 Fällen.

Das im SOEP ursprünglich sechsstufig erfragte Job-Anforderungsniveau wird zunächst aufgrund der Vorüberlegungen von Büchel und Weißhuhn (1997) in drei Kategorien aggregiert: „keine besondere Qualifikation erforderlich", „abgeschlossene Berufsausbildung oder besondere Lehrgänge/Kurse erforderlich" sowie „Hochschulabschluss erforderlich". Es werden die westdeutschen Teilstichproben der Wellen 1984, 1989, 1991, 1993, 1995, 1997, 1998 und 2000 des SOEP in gepoolter Form verwendet. In diesen Jahren wurden Arbeitsplatz-Informationen für alle erwerbstätigen Befragungspersonen ungefiltert erfasst, wohingegen in den übrigen Wellen entsprechende Informationen nur von denjenigen Personen erfragt wurden, die für das Befragungsjahr oder das Vorjahr eine berufliche Veränderung berichten. Die Generierung beschränkt sich auf abhängig Erwerbstätige[8], die sich nicht in einer Vollzeit-Ausbildung befinden. Für das Generierungs-Verfahren stehen schließlich 33.464 Erwerbstätige zur Verfügung.

Das Ergebnis der empirischen Umsetzung dieses Konzepts ist für alle drei Varianten in Tabelle 1 dokumentiert. Dabei wird bereits erkennbar, dass bei Arbeitern und Beamten die Information zur beruflichen Stellung bereits ausreicht, um das dazugehörige Job-Anforderungsniveau hinreichend valide zu indizieren. Dies impliziert, dass beispielsweise Studien zur Fehlqualifikation von Facharbeitern (von Henninges 1996), die sich allein auf

---

6 Wir nutzen den ISCO68 anstelle des ISCO88, da letzterer Code bislang lediglich in drei Wellen des SOEP (mit einer ungefilterten Erhebung des vollständigen Samples; vgl. unten) verfügbar ist. Eine Umstellung des Zuordnungsschemas auf den aktuelleren ISCO88-Code bietet sich erst zu einem Zeitpunkt an, wo das SOEP deutlich mehr Erhebungswellen mit ISCO88-Angaben bereit hält.

7 Die Kodierung des ISCO68 erfolgt anhand der Klarschriftangaben der Respondenten. Sind diese Angaben zu ungenau, greifen die Vercoder in den Erhebungsinstituten häufig ebenfalls auf die Angaben zur beruflichen Stellung zurück. Es ist daher unvermeidlich, dass in Einzelfällen die Angabe zur beruflichen Stellung keine zusätzliche Information zur ausgeübten Tätigkeit liefert. Wir gehen aber davon aus, dass die Leistungsfähigkeit unseres Konzepts durch diese Situation nicht nennenswert beeinträchtigt wird.

8 Selbstständige bleiben unberücksichtigt, da sich deren Job-Anforderungsniveau aufgrund der großen Heterogenität der Anforderungsniveaus häufig nicht eindeutig bestimmen lässt.

die berufliche Position stützen, ausbildungsinadäquate Erwerbstätigkeit hinreichend genau erfassen konnten.

Tabelle 1: Dokumentation der Generierung eines Proxys zum Job-Anforderungsniveau aus den Angaben zum ausgeübten Beruf und zur beruflicher Stellung

Generiertes Job-Anforderungsniveau	Berufliche Stellung	Ausgeübter Beruf (ISCO68) Variante I (Fallzahl >=5)	Ausgeübter Beruf (ISCO68) Variante II (Fallzahl >=10)	Ausgeübter Beruf (ISCO68) Variante III (Fallzahl >=20)
keine Ausbildung erforderlich	Un- und angelernte Arbeiter	alle	alle	alle
	Angestellte mit einfachen Tätigkeiten	322, 331, 339, 342, 370, 380, 395, 399, 432, 452, 532, 540, 551, 552, 560, 589, 795, 842, 949, 971, 985, 999	322, 331, 339, 342, 370, 380, 395, 399, 432, 452, 532, 540, 551, 552, 560, 589, 971, 985, 999	322, 331, 339, 380, 395, 399, 432, 452, 532, 540, 551, 552, 971, 985
Abgeschlossene Berufsausbildung erforderlich	Facharbeiter, Vorarbeiter, Meister	alle	alle	alle
	Angestellte mit einfachen Tätigkeiten	14, 32, 54, 68, 71, 72, 75, 76, 134, 180, 193, 310, 321, 259, 359, 391, 393, 394, 422, 441, 451, 531, 570, 581, 582, 599, 627, 700, 773, 776, 791, 841, 843, 849, 922	14, 54, 71, 72, 75, 76, 134, 193, 310, 321, 391, 393, 394, 451, 531, 570, 582, 599, 627, 776, 791, 849	71, 75, 134, 193, 310, 321, 391, 393, 394, 451, 531, 570, 599, 627
	Angestellte mit qualifizierten Tätigkeiten	14, 32, 33, 34, 35, 38, 39, 54, 68, 71, 72, 73, 75, 76, 77, 81, 83, 84, 129, 134, 135, 159, 162, 163, 172, 174, 180, 191, 193, 194, 199, 211, 212, 219, 300, 310, 321, 322, 331, 339, 342, 359, 391, 392, 393, 394, 395, 399, 400, 421, 422, 431, 432, 441, 442, 451, 500, 520, 531, 532, 540, 551, 552, 570, 589, 599, 627, 700, 773, 776, 795, 833, 841, 842, 843, 849, 852, 855, 861, 871, 892, 922, 925, 971, 985	14, 32, 33, 34, 35, 39, 54, 68, 71, 72, 75, 76, 83, 84, 134, 159, 162, 163, 180, 191, 193, 194, 199, 219, 300, 310, 321, 322, 331, 339, 342, 359, 391, 392, 393, 394, 395, 399, 400, 421, 422, 431, 432, 441, 442, 451, 520, 531, 532, 540, 551, 570, 589, 599, 700, 773, 841, 842, 843, 849, 852, 855, 985	14, 32, 33, 34, 35, 39, 54, 68, 71, 72, 76, 83, 84, 134, 159, 162, 180, 193, 194, 219, 310, 321, 322, 331, 339, 342, 359, 391, 392, 393, 394, 400, 421, 422, 431, 432, 441, 442, 451, 531, 532, 589, 599, 700, 842, 852, 985
	Angestellte mit hochqualifizierten Tätigkeiten	14, 33, 34, 35, 39, 54, 71, 72, 75, 76, 134, 199, 211, 300, 310, 321, 331, 339, 342, 359, 391, 392, 393, 394, 400, 421, 422, 431, 432, 441, 442, 451, 500, 520, 531, 599, 700, 773, 841, 842, 843, 855, 927	14, 33, 34, 35, 39, 54, 71, 76, 134, 211, 300, 310, 321, 331, 339, 342, 359, 391, 392, 393, 394, 400, 421, 422, 431, 432, 441, 442, 451, 500, 700	33, 34, 35, 39, 54, 71, 211, 300, 310, 331, 339, 342, 359, 391, 393, 394, 400, 421, 422, 431, 432, 442, 451, 500, 700
	Beamte im einfachen oder mittleren Dienst	alle	alle	alle
Abgeschlossener Hochschulabschuss erforderlich	Angestellte mit qualifizierten Tätigkeiten	21, 22, 23, 24, 28, 61, 67, 90, 110, 132, 133, 141, 171, 195	21, 22, 23, 24, 90, 110, 132, 133, 171	22, 23, 24, 90, 132, 133
	Angestellte mit hochqualifizierten Tätigkeiten	11, 12, 13, 21, 22, 23, 24, 25, 28, 29, 31, 32, 51, 52, 61, 63, 65, 67, 81, 82, 83, 84, 90, 110, 121, 129, 131, 132, 133, 139, 141, 149, 159, 162, 171, 174, 180, 191, 192, 193, 194, 195, 202, 212, 219, 581	11, 12, 21, 22, 23, 24, 25, 28, 29, 51, 52, 61, 63, 65, 67, 81, 82, 83, 84, 90, 110, 121, 129, 131, 132, 133, 139, 141, 149, 159, 162, 171, 174, 180, 191, 192, 193, 194, 195, 212, 219	11, 21, 22, 23, 24, 25, 28, 29, 61, 67, 83, 84, 90, 110, 129, 131, 132, 133, 159, 162, 193, 194, 212, 219
	Beamte im gehobenen oder höheren Dienst	alle	alle	alle

3. Validitäts-Tests

3.1 Untersuchungsansatz

Die Validierung des beschriebenen Zuordnungsschemas erfolgt anhand eines zweiten Datensatzes. Dieser muss zunächst die beiden Input-Variablen „ausgeübter Beruf“ und „berufliche Stellung“ enthalten. Aus diesen wird mit dem entwickelten Verfahren ein Proxy für das Job-Anforderungsniveau generiert. Diese Variable wird dann mit dem direkt erfragten Job-Anforderungsniveau verglichen. Die Verfügbarkeit der erfragten Arbeitsanforderungen ist somit eine weitere Bedingung, die ein geeigneter Test-Datensatz erfüllen muss.

Für den Validitäts-Test eignet sich die aktuellste Befragung der vom Max-Planck-Institut für Bildungsforschung erhobenen Westdeutschen Lebensverlaufsstudie. Diese enthält retrospektiv erfragte Lebensverlaufsdaten – insbesondere Ausbildungs- und Berufsverläufe – verschiedener Geburtskohorten.[9] Bei der 1998 realisierten Erhebung der Geburtskohorten 1964 und 1971 wurden die Befragungspersonen auch nach dem Anforderungsniveau ihres letzten bzw. – falls erwerbstätig – aktuellen Arbeitsplatzes gefragt. Dabei wurden ähnliche Antwortkategorien wie im SOEP verwendet: Die Befragten gaben an, ob für ihre Berufstätigkeit keine abgeschlossene Berufsausbildung, eine nicht-akademische Berufsausbildung oder ein Hochschulabschluss erforderlich war.

Die Validierung des Generierungsschemas erfolgt in zwei Schritten. In einem ersten Schritt wird die Häufigkeit der erfragten ausbildungsinadäquaten Erwerbstätigkeit mit der Häufigkeit der anhand des generierten Proxys zum Job-Anforderungsniveau gemessenen ausbildungsinadäquaten Erwerbstätigkeit verglichen. Die Höhe der Abweichung zwischen diesen beiden Werten wird als Indikator für die Leistungsfähigkeit des Zuordnungsschemas interpretiert. Ferner wird untersucht, inwiefern der generierte Proxy zum Job-Anforderungsniveau die individuelle Einschätzung der einzelnen Respondenten hinsichtlich der Ausbildungsadäquanz ihrer Tätigkeit korrekt repliziert.

---

9 Für Informationen zu dieser Datenbasis vgl. Corsten und Hillmert (2001). Hinweis gemäß den Vorgaben der Projektleitung: "Die hier verwendeten Lebensverlaufsdaten wurden im Rahmen des Forschungsprojekts 'Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland' erhoben. Dieses Projekt wird vom Max-Planck-Institut für Bildungsforschung in Berlin (Forschungsbereich Bildung, Arbeit und gesellschaftliche Entwicklung) in Kooperation mit dem Institut für Arbeitsmarkt- und Berufsforschung in Nürnberg durchgeführt, welches hierfür mit Mitteln des Europäischen Sozialfonds unterstützt wird. Die Daten (Stand: 15. 4. 2002) befinden sich noch im Prozess der Prüfung und Edition. Die empirischen Ergebnisse sind daher als vorläufig zu betrachten. Die Daten werden nach Abschluss der Prüf- und Editionsarbeiten auf Antrag anderen Forschern zugänglich sein."

In einem zweiten Schritt der Validierung wird untersucht, inwiefern klassische multivariate Auswertungen der Fehlqualifikations-Forschung, in denen der generierte Proxy des Job-Anforderungsniveaus verwendet wird, ähnliche Ergebnisse produzieren wie Analysen, die auf dem erfragten Job-Anforderungsniveau basieren. Dieser Test erfolgt in zwei Teilschritten. Zunächst wird die Ausbildungsadäquanz als abhängige Variable verwendet; wir ermitteln mit einem Logit-Modell die Wahrscheinlichkeit, unterwertig anstelle von ausbildungsadäquat beschäftigt zu sein. In einem zweiten Teilschritt wird die Ausbildungsadäquanz als erklärende Variable in die Analyse aufgenommen. Als Anwendungsbeispiel ermitteln wir mit einer OLS-Regression den Einkommenseffekt einer unterwertigen Beschäftigung. In beiden Anwendungsbeispielen nutzen wir die aus der Literatur etablierten Kovariatensätze (vgl. Büchel 1998).

3.2 Test-Ergebnisse

3.2.1 Abgleich des erfragten mit dem generierten Job-Anforderungsniveau

Tabelle 2 gibt die Fallzahlentwicklung sowie die Häufigkeit der unterwertigen Beschäftigung in Abhängigkeit vom verwendeten Messkonzept wieder. Der obere Teil der Tabelle (Messkonzept A) beinhaltet die Resultate für das Verfahren, das sich auf das erfragte Job-Anforderungsniveau bezieht, der untere Teil (Messkonzept B) die Resultate für das Verfahren, bei dem das Arbeitsanforderungsniveau aus den Informationen zu ausgeübtem Beruf und beruflicher Stellung generiert wurde. Bei Messkonzept B werden die Ergebnisse für alle drei in Abschnitt 2 erwähnten Varianten ausgewiesen.

Nach Ausschluss von Selbstständigen, Auszubildenden und Studierenden sowie Erwerbstätigen ohne Berufsausbildung umfasst das Sample 2.158 Personen. Soll das erfragte Job-Anforderungsniveau zur Erzeugung der Mismatch-Variablen genutzt werden, ergibt sich folgende weitere Fallzahlentwicklung: 24 Personen können aufgrund fehlender Angaben zum Job-Anforderungsniveau nicht berücksichtigt werden. Von den verbleibenden 2.134 Personen verfügen 179 Erwerbstätige gemäß dem von Büchel und Weißhuhn (1997) entwickelten Kategorisierungs-Schema über unplausible Kombinationen von formalem Qualifikationsniveau, beruflicher Stellung und Job-Anforderungsniveau und werden entsprechend ebenfalls ausgeschlossen. Schließlich verbleiben 1.955 Befragungspersonen bzw. 90,6% der ursprünglichen Population mit einem gültigen Wert zur Ausbildungsadäquanz ihrer Beschäftigung. Sie bilden damit das Ausgangssample für unsere Analysen.

Tabelle 2: Fallzahlentwicklung und Häufigkeit von ausbildungsinadäquater Erwerbstätigkeit in Abhängigkeit des Messkonzepts

	N	Prozent
Messkonzept A: erfragtes Job-Anforderungsniveau		
Alle Erwerbstätige (ohne Selbstständige, Personen in Ausbildung, Erwerbstätige ohne Berufsausbildung)	2.158	100,0
mit gültigen Angaben zum erfragten Anforderungsniveau	2.134	98,9
mit plausibler Kombination von Qualifikationsniveau, beruflicher Stellung und Job-Anforderungsniveau	1.955	90,6
davon ausbildungsinadäquat erwerbstätig		10,0
Messkonzept B: generierter Proxy zum Job-Anforderungsniveau		
Alle Erwerbstätige (ohne Selbstständige, Personen in Ausbildung, Erwerbstätige ohne Berufsausbildung)	2.158	100,0
mit gültigen Angaben zur beruflichen Stellung und Berufstätigkeit	2.126	98,5
Variante I:		
Kombination von ausgeübtem Beruf und beruflicher Stellung: Fallzahl mindestens 5	1.977	91,6
davon ausbildungsinadäquat erwerbstätig		11,1
Variante II:		
Kombination von ausgeübtem Beruf und beruflicher Stellung: Fallzahl mindestens 10	1.926	89,2
davon ausbildungsinadäquat erwerbstätig		11,2
Variante III:		
Kombination von ausgeübtem Beruf und beruflicher Stellung: Fallzahl mindestens 20	1.843	85,4
davon ausbildungsinadäquat erwerbstätig		11,4

Datenbasis: Westdeutsche Lebensverlaufsstudie, Kohorten 1964 und 1971 (Beobachtungsjahr: 1998); eigene Auswertung.

Soll die Ausbildungsadäquanz der Beschäftigung mit Hilfe des generierten Proxys zum Job-Anforderungsniveau festgestellt werden, so ergibt sich eine andere Fallzahlentwicklung. Das Startsample umfasst wiederum 2.158 Personen. 32 Personen müssen zunächst aufgrund fehlender Angaben zu ausgeübtem Beruf oder beruflicher Stellung ausgeschlos-

sen werden. Bei der Anwendung der Variante II des Zuordnungsschemas können weiteren 200 Personen (Variante I: 149 Personen; Variante III: 283 Personen) aufgrund zu geringer Fallzahlen hinsichtlich der Kombination von beruflicher Stellung und ausgeübtem Beruf kein Proxy zum Job-Anforderungsniveau zugeordnet werden. Es verbleiben 89,2 Prozent des Startsamples (Variante I: 91,2; Variante III: 85,4). Das subjektive Messkonzept und die Varianten I und II des oben entwickelten Messkonzepts führen somit zu einem ähnlich hohen Ausschluss aufgrund ungültiger Werte; bei Variante III fällt die Fallzahlreduzierung etwas stärker aus.

Die Berechnung des Anteils der ausbildungsinadäquat Erwerbstätigen an allen Erwerbstätigen anhand des erfragten Job-Anforderungsniveaus ergibt einen Wert von 10,0 Prozent. Die Anwendung des generierten Proxys zum Job-Anforderungsniveau der Variante II des Zuordnungsschemas ergibt einen leicht höheren Anteil von 11,2 % (Variante I: 11,1 %, Variante III: 11,4 %). Die Differenz von 1,1 bis 1,4 Prozentpunkten mag zunächst als nicht unerhebliche Abweichung erscheinen. Sie ist allerdings vor dem Hintergrund zu bewerten, dass die Fehlqualifikations-Forschung – anders als beispielsweise die Arbeitslosen-Forschung – über keine standardisierten Messkonzepte verfügt. Entsprechend groß ist die Bandbreite der berichteten Anteile an unterwertiger Beschäftigung selbst in Fällen, in denen identische Datenbasen genutzt wurden (vgl. hierzu ausführlich Büchel 1998: 66ff.). Die Disziplin hat auf diese Problematik dahingehend reagiert, dass sie weniger Niveaus als Strukturmuster von unterwertiger Beschäftigung interpretiert. Entsprechend gilt den Ergebnissen der nachfolgenden multivariaten Analysen unser besonderes Augenmerk.

Zuvor gibt jedoch Tabelle 3 darüber Aufschluss, wie gut das Zuordnungsschema die individuelle subjektive Einschätzung der Befragungspersonen hinsichtlich der Ausbildungsadäquanz ihrer Tätigkeit replizieren kann: Aus den in Tabelle 2 ausgewiesenen Ergebnissen ähnlich hoher Aggregatwerte lässt sich noch nicht schließen, dass mit unserem Zuordnungsschema auch auf der Ebene der einzelnen Befragten eine gute Übereinstimmung von erfragtem und generiertem Job-Anforderungsniveau erzielt werden konnte.[10] Die Ergebnisse belegen eine überaus gute Zuordnungsleistung. Bei 98,3 % der befragten Personen stimmt die subjektiv empfundene mit der anhand des Zuordnungsschemas (Variante II) generierten Ausbildungsadäquanz überein. Nur 1,4 % der Befragungspersonen werden fälschlicherweise durch ein zu geringes generiertes Job-Anforderungsniveau als unterwertig beschäftigt eingestuft, obwohl sie berichten, ausbildungsadäquat erwerbstätig zu sein. Das Risiko einer Überschätzung des Job-Anforderungsniveaus ist noch geringer: Ganze 0,3 % der Personen werden fälschlicherweise durch ein zu hoch

10 Bei dieser Analyse können nur Personen berücksichtigt werden, die sowohl gültige Angaben zur subjektiven als auch zur generierten Ausbildungsadäquanz aufweisen.

generiertes Job-Anforderungsniveau als ausbildungsadäquat eingestuft, obwohl sie nach eigener Einschätzung unterwertig beschäftigt sind. Diese Fehlzuordnungen erscheinen als geringfügig.

Tabelle 3: Abgleich der „generierten" Ausbildungsadäquanz (Konzept B) mit der „erfragten" Ausbildungsadäquanz (Konzept A)

	Variante I Prozent	Variante II Prozent	Variante III Prozent
„Generierte" Ausbildungsadäquanz = „erfragte" Ausbildungsadäquanz	98,2	98,3	98,4
„Generierte" Ausbildungsadäquanz = AAB & „erfragte" Ausbildungsadäquanz = UWB	0,4	0,3	0,2
„Generierte" Ausbildungsadäquanz = UWB & „erfragte" Ausbildungsadäquanz = AAB	1,4	1,4	1,4
Insgesamt	100,0	100,0	100,0
N	1.807	1.760	1.687

Datenbasis: Westdeutsche Lebensverlaufsstudie, Kohorten 1964 und 1971 (Befragungsjahr: 1998); eigene Auswertung.

AAB = Ausbildungsadäquat beschäftigt, UWB = Unterwertig beschäftigt

3.2.2 Anwendungsbeispiel: Determinanten einer unterwertigen Erwerbstätigkeit

Tabelle 4 gibt die Ergebnisse dreier Logit-Modelle wieder.[11] Als abhängige Variable wird jeweils die Ausbildungsadäquanz der Beschäftigung verwendet. Im ersten Modell wird diese anhand des erfragten Job-Anforderungsniveaus gemessen, im zweiten anhand des

11 Im Interesse einer übersichtlichen Ergebnisdarstellung verfolgen wir nur eine Zuordnungs-Variante bei den multivariaten Analysen. Bei der Auswahl einer bestimmten Variante verwerfen wir zunächst Variante III, da hier die Fallzahlverluste unnötig groß erscheinen. Danach geben wir Variante II den Vorzug gegenüber Variante I, da sie – bei nur geringen Fallzahlverlusten – eine höhere Robustheit gegen atypische Kombinationen von beruflicher Stellung, ausgeübtem Beruf und Job-Anforderungsniveau erwarten lässt als Variante I. Die (nicht dokumentierten) Ergebnisse der Auswertungen, die auf Variante I und III basieren, weichen nicht wesentlich von den ausgewiesenen Ergebnissen ab.

generierten Proxys zum Job-Anforderungsniveau[12].

Tabelle 4: Determinanten einer ausbildungsinadäquaten Erwerbstätigkeit (Logistische Regression)

	Messkonzept: Erfragtes Job-Anforderungsniveau	**Messkonzept: Generierter Proxy**	**Gepoolte Messkonzepte**
	Koeffizient (Std.-fehler)	Koeffizient (Std.-fehler)	Koeffizient (Std.-fehler)
(Konstante)	0,6602	0,4850	0,5908
	(0,3393)	(0,3325)	(0,3251)
Mann	-1,4079**	-1,1535**	-1,2727**
	(0,2086)	(0,2039)	(0,2104)
1971er Kohorte (vs. 1964er Kohorte)	-0,3286	-0,3499	-0,3395
	(0,1875)	(0,1892)	(0,1757)
Ausländer	0,1395	-0,0765	0,0395
	(0,4893)	(0,5059)	(0,4883)
Schulabschluss (Hauptschulabschluss)			
Realschulabschluss	-1,2846**	-1,0500**	-1,1550**
	(0,2221)	(0,2156)	(0,2043)
Abitur	-2,1988**	-1,6898**	-1,9545**
	(0,5347)	(0,5543)	(0,4773)
Ausbildungsabschluss (Lehre / Berufsfachschule)			
Fachhochschulabschluss	1,8493**	2,5920**	2,1395**
	(0,3399)	(0,3776)	(0,3524)
Hochschulabschluss	1,1046**	1,8496**	1,4045**
	(0,3064)	(0,3510)	(0,3111)
Relative Berufserfahrung	-0,1853	-0,0659	-0,1371
	(0,3648)	(0,3613)	(0,3413)
Betriebszugehörigkeitsdauer (in Jahren)	-0,1183**	-0,1001**	-0,1081**
	(0,0310)	(0,0288)	(0,0283)
Angestellter (vs. Arbeiter)	-2,1901**	-3,2734**	-2,6714**
	(0,2402)	(0,2863)	(0,2352)
Öffentlicher Dienst (vs. Privatwirtschaft)	-0,9154**	-0,6833*	-0,8066*
	(0,3185)	(0,3180)	(0,3258)
Betriebsgröße (logarithmiert)	0,1014*	0,1214**	0,1113**
	(0,0420)	(0,0394)	(0,0374)
Generierter Proxy (vs. erfragtes Anforderungsniveau)			0,0130
			(0,0592)

12 Das erste Modell berücksichtigt Personen mit gültigen Angaben zum erfragten Job-Anforderungsniveau, wogegen das zweite Modell Personen mit gültigen Angaben zum generierten Job-Anforderungsniveau einbezieht. Die Untersuchungspopulationen der beiden Modelle sind somit nicht identisch.

Signifikanzniveau: ** = p < 0,01; * = p < 0,05.			
Abhängige Variable: 1 = unterwertig beschäftigt; 0 = ausbildungsadäquat beschäftigt.			
Mittelwert der abhängigen Variablen	10,4	11,9	11,2
Fallzahl	1.751	1.725	3.476
Log-Likelihood	-450,07**	-453,55**	-912,52**
Pseudo-R^2	0,2325	0,2834	0,2523

Datenbasis: Westdeutsche Lebensverlaufsstudie, Kohorten 1964 und 1971 (Beobachtungsjahr: 1998); eigene Auswertung.

Die Ergebnisstruktur des ersten Logit-Modells gibt die aus der Literatur bekannten Befunde wieder. Wir verzichten an dieser Stelle auf eine inhaltliche Würdigung der Ergebnisse (vgl. hierzu zum Beispiel Büchel 1998). Ein Vergleich der Ergebnisse des ersten Modells mit jenen des zweiten macht deutlich, dass die Verwendung des generierten Proxys zum Job-Anforderungsniveau anstelle des erfragten Job-Anforderungsniveaus eine recht ähnliche Ergebnis-, insbesondere Signifikanzstruktur hervorbringt. Dieses Ergebnis legt nahe, dass Fehlqualifikations-Studien bei Anwendung des oben entwickelten Zuordnungsschemas zu gleichen Schlüssen gelangen wie solche, die auf Datensätze mit Informationen zum erfragten Job-Anforderungsniveau zurückgreifen können.[13]

Das dritte Modell in Tabelle 4 basiert auf den gepoolten Fallzahlen der ersten beiden Modelle. Bei dieser Analyse sind alle Befragungspersonen mit gültigen Angaben zum erfragten Job-Anforderungsniveau und zum generierten Proxy des Job-Anforderungsniveau doppelt vertreten[14]. Eine zusätzlich zum bestehenden Kovariatensatz aufgenommene Dummy-Variable indiziert, ob die Ausbildungsadäquanz der Untersuchungspersonen anhand des erfragten oder aber des generierten Proxys zum Job-Anforderungsniveau ermittelt wurde.

Wie aufgrund der deskriptiven Ergebnislage erwartet, wird bei Verwendung des generierten Proxys zum Job-Anforderungsniveau anstelle der erfragten Information das Vorliegen einer ausbildungsinadäquaten Erwerbstätigkeit etwas zu oft konstatiert. Allerdings ist der Effekt nur sehr gering und nicht signifikant. Das Zuordnungsschema erweist sich damit bei dieser Standard-Auswertung der Überqualifikations-Forschung als sehr robust.

13 Die Höhe der geschätzten Koeffizienten ist dagegen – wie allgemein üblich – nur zurückhaltend zu interpretieren.

14 Auf das sich daraus ergebende methodologische Problem fehlender Unabhängigkeit der Untersuchungseinheiten wird bei der Wahl des statistischen Auswertungsverfahrens in angemessener Form reagiert.

3.2.3 Anwendungsbeispiel: Einkommenseffekte einer unterwertigen Erwerbstätigkeit

In unserem zweiten Anwendungsbeispiel analysieren wir die negativen Einkommenseffekte von ausbildungsinadäquater Erwerbstätigkeit. Im Gegensatz zum ersten Anwendungsbeispiel geht die Ausbildungsadäquanz nun als unabhängige Variable in die Analyse ein. Analog zum vorhergehenden Anwendungsbeispiel wird in einem ersten Modell die Ausbildungsadäquanz anhand des erfragten Job-Anforderungsniveaus gemessen, wogegen in einem zweiten Modell die Ausbildungsadäquanz anhand des generierten Proxys zum Job-Anforderungsniveau indiziert wird. Wie im vorhergehenden Analyseschritt zeigt ein Vergleich der Ergebnisse beider Modelle recht ähnliche Ergebnisse (Tabelle 5).

Tabelle 5: Einkommenseffekte von ausbildungsinadäquater Erwerbstätigkeit (OLS, verschiedene Messkonzepte)

	Messkonzept: Erfragtes Job-Anforderungsniveau	**Messkonzept: Generierter Proxy**	**Gepoolte Messkonzepte**
	Koeffizient (Std.-fehler)	Koeffizient (Std.-fehler)	Koeffizient (Std.-fehler)
(Konstante)	2,1179**	2,1201**	2,1185**
	(0,0643)	(0,0619)	(0,0747)
Mann	0,1972**	0,2102**	0,2036**
	(0,0157)	(0,0156)	(0,0155)
1971er Kohorte (vs 1964er Kohorte)	-0,1186**	-0,0940**	-0,1063**
	(0,0180)	(0,0176)	(0,0194)
Ausländer	0,0149	0,0205	0,0179
	(0,0411)	(0,0399)	(0,0331)
Bildungsdauer (in Jahren)	0,0441**	0,0436**	0,0439**
	(0,0036)	(0,0036)	(0,0043)
Berufserfahrung (in Jahren)	0,0513**	0,0513**	0,0514**
	(0,0062)	(0,0061)	(0,0066)
Berufserfahrung**2/100	-0,2426**	-0,2429**	-0,2429**
	(0,0372)	(0,0364)	(0,0378)
Betriebszugehörigkeitsdauer	0,0011	0,0024	0,0017
(in Jahren)	(0,0024)	(0,0023)	(0,0024)
Angestellter (vs. Arbeiter)	0,0255	0,0369	0,0302
	(0,0199)	(0,0196)	(0,0186)
Beamter (vs. Arbeiter)	-0,1305**	-0,1180**	-0,1248**
	(0,0366)	(0,0347)	(0,0340)
Öffentlicher Dienst	-0,0002	0,0072	0,0033
(vs. Privatwirtschaft)	(0,0197)	(0,0196)	(0,0200)
Betriebsgröße (logarithmiert)	0,0325**	0,0340**	0,0332**
	(0,0035)	(0,0034)	(0,0034)
Ausbildungsinadäquat beschäftigt	0,1151**	0,0732**	0,0932**
	(0,0251)	(0,0237)	(0,0226)

Generierter Proxy (vs. erfragtes Anforderungsniveau)			0,0021 (0,0032)
Signifikanzniveau: **= p<0,01; *= p<0,05.			
Abhängige Variable: log(Bruttostundenlohn)			
Mittelwert der abhängigen Variablen:	3,165	3,167	3,166
Fallzahl	1.661	1.639	3.300
R^2	0,3179	0,3254	0,3206

Datenbasis: Westdeutsche Lebensverlaufsstudie, Kohorten 1964 und 1971 (Beobachtungsjahr: 1998); eigene Auswertung.

Das dritte Modell beruht wiederum auf dem gepoolten Datensatz, wobei eine zusätzliche Dummy-Variable das verwendete Messkonzept der Ausbildungsadäquanz (basierend auf erfragtem vs. generiertem Proxy zum Job-Anforderungsniveau) indiziert. Der für diese Variable geschätzte Parameter ist wie im vorhergehenden Analyseschritt sehr gering und nicht signifikant. Unser Zuordnungsschema erweist sich damit auch hinsichtlich dieser Standard-Auswertung der Überqualifikations-Forschung als robust.

4. Fazit

Für die Bestimmung der Ausbildungsadäquanz der Beschäftigung werden Informationen zur formalen Qualifikation der Erwerbstätigen sowie zum Job-Anforderungsniveau benötigt. Während erstere Angabe in allen gängigen Datensätzen, die der mikrodatenbezogenen Arbeitsmarktforschung zur Verfügung stehen, enthalten ist, gilt dies nicht für das Job-Anforderungsniveau. Dies hemmte bislang die Entwicklung der deutschen Fehlqualifikations-Forschung erheblich. Das von uns vorgestellte Verfahren, ein Proxy zum Job-Anforderungsniveau aus den Informationen zum ausgeübten Beruf und zur beruflichen Stellung (die in den allermeisten Datensätzen verfügbar sind) zu generieren, erweist sich als leistungsfähig. Multivariate Anwendungsbeispiele führen unabhängig vom Messkonzept zu sehr ähnlichen Ergebnissen, wobei sich die Ähnlichkeit allerdings mehr auf die Signifikanzstruktur als auf die Höhe der geschätzten Parameter bezieht. Diese Einschränkung dürfte jedoch kaum auf ein besonderes Problem unseres Zuordnungsschemas hindeuten, wenn man bedenkt, dass die Ergebnisse multivariater Analysen aufgrund von Unschärfen in Befragungsdaten ohnehin vorzugsweise hinsichtlich ihrer Signifikanzstruktur und weniger in Bezug auf die Höhe der einzelnen geschätzten Parameter interpretiert werden sollten (vgl. Büchel 2000). Diese allgemein gültigen Einschränkungen berücksichtigend erhält die deutsche Fehlqualifikations-Forschung somit durch unser neues Tool die Möglichkeit, sich von den wenigen Datensätzen zu lösen, die das Job-Anforderungsniveau direkt erfragen. Dies könnte eine weitere Entfaltung dieser überaus dynamischen Disziplin begünstigen.

Korrespondenzadresse

Matthias Pollmann-Schult & Felix Büchel
Max-Planck-Institut für Bildungsforschung
Lentzeallee 94
14195 Berlin
Emails: pollmann@mpib-berlin.mpg.de; buechel@mpib-berlin.mpg.de

Literatur

Büchel, F., 1998: Zuviel gelernt? Ausbildungsinadäquate Erwerbstätigkeit in Deutschland. Bielefeld: Bertelsmann Verlag.

Büchel, F., 2000: Institutseffekte bei Befragungen - Auswirkungen auf Datenqualität und Analyseergebnisse. Allgemeines Statistisches Archiv 84: 417-446.

Büchel, F., 2001: Overqualification: reasons, measurement issues and typological affinity to unemployment. S. 453-560 in: P. Descy/M. Tessaring (Hrsg.), Training in Europe Second report on vocational training research in Europe 2000: background report. CEDEFOP Reference series. Luxembourg: Office for Official Publications of the European Communities.

Büchel, F./Weißhuhn, G., 1997: Ausbildungsinadäquate Beschäftigung bei Absolventen des Bildungssystems: Berichterstattung zur Struktur und Entwicklung unterwertiger Beschäftigung in West- und Ostdeutschland. Berlin: Duncker & Humblot.

Corsten, M./Steffen H., 2001: Qualifikation, Berufseinstieg und Arbeitsmarktverhalten unter Bedingungen erhöhter Konkurrenz. Was prägt Bildungs- und Erwerbsverläufe in den achtziger Jahren? Berlin: Max-Planck-Institut für Bildungsforschung, Arbeitspapier Nr. 1 des Projekts 'Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland'.

Hartog, J., 2000: Over-education and earnings: where are we, where should we go? Economics of Education Review 19: 131-147.

Henninges, H. von, 1996: Steigende Qualifikation im Arbeiterbereich? Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 29: 73-92.

Konietzka, D., 1999: Die Verberuflichung von Marktchancen. Die Bedeutung des Ausbildungsberufes für die Plazierung im Arbeitsmarkt. Zeitschrift für Soziologie 28: 379-400.

Plicht, H./Schober, K./Schreyer, F., 1994: Zur Ausbildungsadäquanz der Beschäftigung von Hochschulabsolventinnen und -absolventen. Versuch einer Quantifizierung anhand der Mikrozensen 1985 bis 1991. Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 27: 177-204.

SOEP Group, 2001: The German Socio-Economic Panel GSOEP after more than 15 years - Overview. Vierteljahrshefte zur Wirtschaftsforschung (Sonderheft) 70 (1): 7-14.

Von der Schwierigkeit, Rechtsfragen Sozialwissenschaftlich zu Untersuchen.

Replik zu Häder/Klein ZUMA-Nachrichten 50: 86 – 111

Arthur Hartmann

Der Artikel „Wie wenig das Recht unser Verhalten regelt."[1] stellt sich als besonders sorgfältige Untersuchung zu einem schwierigen Forschungsthema vor. Der Untertitel: „Methodische Innovationen und erste Befunde zu einem bisher nicht untersuchten Thema." macht deutlich, dass Häder/Klein den methodischen Aspekten ihrer Arbeit das gleiche Gewicht beimessen wie den inhaltlichen Befunden. Bei näherer Analyse ergeben sich aber m.E. wichtige Kritikpunkte, die im Hinblick darauf, dass es sich um den Erstbericht zu einem umfangreichen, von der Volkswagenstiftung geförderten Projekt handelt, möglichst frühzeitig zur Diskussion gestellt werden sollen. Diese Kritik betrifft den methodischen Zugang, die Schlussfolgerungen und Probleme der Umsetzung von Rechtsfragen in das Forschungsprogramm.

Die Untersuchung umfasste 3463 standardisierte Telefoninterviews, wobei die Befragten 19 Sachverhalte/Fälle bewerten sollten, die größtenteils der Rechtsprechung zur Notwehr entnommen wurden. Die Antwortmöglichkeiten beschränkten sich überwiegend auf die Kategorien: „Halten Sie das Verhalten des XY für gerechtfertigt bzw. nicht für gerechtfertigt".[2] Die Originalfälle wurden gekürzt und z.T. verändert, wobei insbesondere Merkmale, die Vorurteile oder sachfremde Emotionen wecken könnten, durch neutrale Merkmale ersetzt oder gestrichen wurden.[3] Leider werden die verwendeten Urteile nicht zitiert.

---

1 ZUMA-Nachrichten 50, Jg. 26, Mai 2002, S. 86-112.

2 Häder/Klein, ZUMA-Nachrichten 50, 86, 91 ff., 107 ff.

3 Häder/Klein, ZUMA-Nachrichten 50, 91 ff.; Im Fall q12 allerdings beleidigt ein Mann einen anderen mit dem Vorwurf, der andere sei an seiner Kriegsgefangenschaft in der Sowjetunion selbst schuld. Dies ist wegen der kontroversen Diskussion über die Verbrechen der Wehrmacht in der Sowjetunion kein neutraler Sachverhalt. Im Fall q18a wird der Ausgangsfall q18, in dem ein Unter-

Als Ziel der Untersuchung nennen Häder/Klein: „Die Projektmitarbeiter versprachen sich von der Befragung Befunde, welche u.U. Prinzipien der Rechtsprechung in Deutschland in Frage stellen."[4] Die Befragten sollten 19 ethisch und dogmatisch schwierige Fälle, die die Prinzipien des Notwehrrechtes berühren und Revisionsgerichte beschäftigt haben, am Telefon spontan entscheiden, ohne sich zuvor mit der Problematik auseinander setzen zu können. Eine solche Vorgehensweise erscheint bei Meinungsumfragen zu Vorlieben beim Kauf von Waschmitteln etc. angebracht. Im Hinblick auf das anspruchsvolle Forschungsziel wäre es jedoch angemessen gewesen, den Befragten einige Zeit zum Nachdenken und zur Auseinandersetzung mit der Thematik einzuräumen.[5] Als Erhebungsmethode wäre deshalb eine schriftliche Befragung vorzugswürdig gewesen. Mit Gruppendiskussionen hätten darüber hinaus nicht nur die Bewertungen, sondern auch die Argumente der Befragten erfasst werden können.

Bei der Darstellung ihrer Vorgehensweise erwähnen Häder/Klein zwar, dass die Reihenfolge der Fragen permutiert wurde, um Reihenfolgeeffekte zu verhindern[6], sie äußern sich aber nicht zu Interviewereffekten. Da die Befragten spontan zu schwierigen Fragen Stellung nehmen mussten, über die sie wohl eher selten nachdenken, könnte der Einfluss der Interviewer erheblich sein. Zwar wurden die Fallschilderungen von einem Tonband eingespielt[7], dennoch verbleiben vielfältige Einflussmöglichkeiten der Interviewer z.B. allein durch zustimmende Kommentierung einer Antwort mit „mhm" oder ablehnende bzw. nachfragende Kommentierung mit „hm".[8] Hinzu kommt, dass die Interviewer anscheinend mit der Position der Wissenschaftler vertraut waren[9], wobei letztere das deutsche Notwehrrecht offenbar ablehnen. Dies kommt an verschiedenen Stellen deutlich zum Ausdruck, insbesondere in der Formulierung „Das deutsche Notwehrrecht befördert eine ‚Schlägermoral'.[10] Es wäre deshalb interessant gewesen zu erfahren, wie die Interviewer/-innen die vorgelegten Fälle bewertet haben und sodann ob eine Korrelation mit den Wertungen der Befragten besteht.

---

nehmer seinen Erpresser erschießt, sogar dahingehend abgewandelt, dass der Unternehmer ein glücklich verheirateter und sozial engagierter Familienvater ist.

4 Häder/Klein, ZUMA-Nachrichten 50, 100.

5 Unter diesem Aspekt waren die Bedingungen im Pretest anscheinend günstiger als in der Hauptuntersuchung; vgl. Häder/Klein, ZUMA-Nachrichten 50, 94, 101.

6 Häder/Klein, ZUMA-Nachrichten 50, 95.

7 Häder/Klein, ZUMA-Nachrichten 50, 95.

8 Vgl. zur Interviewerbias etwa Kromrey, Empirische Sozialforschung, 8. Aufl., 1998, 341 ff., 344 ff.

9 Dies ergibt sich aus dem Hinweis, dass bei dem **Pretest** auch Interviewer mitwirken konnten, die **nicht** mit den inhaltlichen Grundlagen der Befragung vertraut waren; Häder/Klein, ZUMA-Nachrichten 50, 96.

10 Häder/Klein, ZUMA-Nachrichten 50, 89 in Hypothese 3.

Hinsichtlich der bisher vorliegenden Untersuchungsergebnisse weisen Häder/Klein auf Bewertungsdifferenzen zwischen Männern und Frauen hin. Diese Befunde müssen aber mit der Einschränkung versehen werden, dass fast ausschließlich Fälle untersucht wurden, in denen Männer sich gegen Männer verteidigen. Denn nur in einem Fall (q1) erscheint eine Frau in der Rolle der Angreiferin und ebenfalls nur in einem Fall (Fall q17) in der Rolle der Notwehr-Übenden.[11] In der Endauswertung wollen Häder/Klein untersuchen, inwieweit die Bewertung von Notwehrsituationen durch Geschlechterrollen bestimmt wird.[12] Dieses Vorhaben hätte jedoch eine Variation der Geschlechtsrollen in den Fällen erfordert.

Durch die vorliegenden Befunde sehen Häder/Klein ihre Skepsis hinsichtlich der Akzeptanz der Rechtsbewährungstheorie bestätigt.[13] Dies folge daraus, dass der mit Frage q19 untersuchte Rechtsbewährungsgedanke nicht mit soziodemografischen Merkmalen der Befragten und nicht mit wertkonservativen Haltungen korreliert sei.[14] Dieser Argumentation kann nicht gefolgt werden. Fall q19 lautet: In der Zeitung ist zu lesen, dass ein Mann, der von fünf jungen Männern zur Herausgabe seiner Geldbörse gedrängt wurde, einen der jungen Männer erschossen hat.[15] Es verwundert nicht, dass nur 8 % der Befragten dieses Verhalten für gerechtfertigt hielten.[16] Allein wegen dieser Schiefe der Antwortverteilung könnten die von Häder/Klein erwarteten Korrelationen nur größere Werte erreichen, wenn die unabhängigen Variablen „soziodemografische Merkmale“ und „Werthaltungen“ ebenfalls sehr schief verteilt wären. Zudem wird unten bei der Diskussion von Rechtsfragen gezeigt werden, dass mit Fall q19 nicht der Rechtsbewährungsgedanke im Sinne des deutschen Rechts gemessen wird. In diesem Zusammenhang ist schließlich bemerkenswert, dass diejenigen Befragten, die selbst schon eine Notwehrsituation erlebt hatten, den Gedanken der Selbstverteidigung überwiegend positiv beurteilten.[17] Häder/Klein erklären diesen Befund mit der Vermutung, das Notwehrrecht wirke

---

11 Häder/Klein, ZUMA-Nachrichten 50, 107 ff.

12 Vgl. Häder/Klein, ZUMA-Nachrichten 50, 104, 106.

13 Häder/Klein, ZUMA-Nachrichten 50, 87, 89 f., 101 ff.; zur Rechtsbewährungstheorie siehe unten die Diskussion der Rechtsfragen. Vgl. auch Schönke/Schröder/Lenkner/Perron, Strafgesetzbuch – Kommentar, 26. A. 2001, § 32 Rn 1, 1a m.w.N.; Marxen, Die sozialethischen Grenzen der Notwehr 1979.

14 Häder/Klein, ZUMA-Nachrichten 50, 104.

15 Häder/Klein, ZUMA-Nachrichten 50, 111.

16 Häder/Klein, ZUMA-Nachrichten 50, 103.

17 Häder/Klein, ZUMA-Nachrichten 50, 104 f.

nicht präventiv[18] i.S. der Rechtsbewährungstheorie, sondern – im Gegenteil – therapeutisch, denn es diene dazu, das eigene Verhalten in einer solchen Situation nachträglich zu legitimieren.[19] Dies ist eine für die Betroffenen eher abwertende Interpretation. Eine alternative Erklärung lautet, dass die Befragten mit eigener Notwehrerfahrung die Erforderlichkeit und Wirksamkeit des Notwehrrechts besonders fundiert bewerten können und ihr abweichendes Votum die generelle Differenz zwischen moralischen Einstellungen und tatsächlichem Verhalten spiegelt. Der Titel des Aufsatzes „Wie wenig das Recht unser Verhalten regelt" findet demnach in den genannten Befunden keinen Rückhalt.

Der dritte Kritikpunkt betrifft die Umsetzung der rechtlichen Fragestellung in das Forschungsprogramm. Häder/Klein geben zunächst einen knappen Überblick zur gesetzlichen Regelung und stellen die beiden Prinzipien vor, auf denen nach der in der juristischen Literatur vorherrschenden, aber nicht unumstrittenen Meinung (sog. h.M.) die rechtliche Regelung beruht. Es sind dies das individuelle Recht, eigene und fremde Rechtsgüter gegen Angriffe zu verteidigen (sog. Schutzprinzip), und der Grundsatz, dass das Recht nicht vor einem rechtswidrigen Angriff zurückzuweichen braucht, sondern in diesem Fall auch von den Bürgern im Interesse der Allgemeinheit verteidigt werden darf (Rechtsbewährungsprinzip).[20] Damit wird die in dem Artikel von Häder/Klein mehrfach genannte „Schneidigkeit" bzw. „Scharfkantigkeit" des Notwehrrechts begründet, die vor allem darin zum Ausdruck kommt, dass der Angegriffene grundsätzlich zu einem Abwehrmittel greifen darf, das den Angriff sicher beendet, selbst wenn das von dem Angriff bedrohte Gut geringere Wertigkeit hat als das von der Notwehr verletzte Gut.[21] Das Wort „grundsätzlich" bedeutet in der Rechtssprache allerdings, dass Ausnahmen und Grenzen bestehen. Die von Häder/Klein gewählte Formulierung: „Es gilt in Deutschland schließlich das Prinzip, dass das Recht niemals dem Unrecht zu weichen braucht"[22] ist in ihrer Absolutheit deshalb falsch.

---

18 Kriminalprävention ist allerdings nicht die einzige und wohl auch nicht die bedeutsamste Legitimation der Rechtsbewährungstheorie; vgl. Schönke/Schröder/Lenckner/Perron a.a.O. § 32 Rn 1a m.w.N. sowie unten die Diskussion der Umsetzung von Rechtsfragen in der Untersuchung.

19 Häder/Klein, ZUMA-Nachrichten 50, 104 f.

20 Vgl. Hierzu Schönke/Schröder/Lenckner a.a.O. § 32 Rn 1, 1a m.w.N. zum Meinungsstand. Das Rechtsbewährungsprinzip bewirkt nicht nur eine Ausdehnung, sondern kann auch zu einer Beschränkung des Notwehrrechts führen; vgl. Amelung, Das Problem der heimlichen Notwehr gegen die erpresserische Androhung kompromittierender Enthüllungen GA 1982, 392 f., 396 f.

21 Häder/Klein, ZUMA-Nachrichten 50, 87 f.; Kühl, Strafrecht Allgemeiner Teil, 3. A. 2001 S. 170 Rn 88.

22 Häder/Klein, ZUMA-Nachrichten 50, 88.

Bedenken bestehen im Rahmen des dritten Kritikpunktes zunächst gegen die Operationalisierung der Fallfragen, da die telefonische Befragung zu einer sehr starken Verkürzung der Fälle nötigte. Dies lässt sich z.B. am Fall q5 zeigen, der lautet: „Vor kurzem wurde jemandem nachts sein auf der Straße geparktes Auto zerkratzt. Deshalb mustert er – vor seinem Hause stehend – vier junge Männer, weil er befürchtet, dass sein Auto erneut zerkratzt werde. Die jungen Männer fühlen sich dadurch provoziert. Deshalb beginnen die jungen Männer mit Steinen zu werfen. Als die Steinewerfer näher kommen, gibt der Angegriffene einige ungezielte Schüsse ab. Ein Schuss trifft einen Steinewerfer und verursacht eine Querschnittslähmung. Halten Sie das Verhalten des Schützen für gerechtfertigt oder nicht?“[23]

Der Fall dürfte BGH NStZ 1986, 357 f. entsprechen.[24] Der in der BGH-Entscheidung wiedergegebene Sachverhalt lautet: „Vier Jugendliche im Alter von 15 bis 19 Jahren waren ihm (dem Angeklagten, Anm. d. Verf.) in kurzem Abstand gefolgt und wollten auf der Straße weitergehen. Der Angeklagte schloss die Haustür auf, ohne jedoch das Schnappschloss zu betätigen, und ließ den Schlüssel stecken. Da ihm etwa ein Jahr zuvor Teile seines auf der Straße abgestellten Leichtkraftrads abhanden gekommen waren, wollte er beobachten[25], was die Jugendlichen vorhatten. Sie wandten sich dem Angeklagten zu, näherten sich ihm und lösten mit der Frage, ‚Was willst Du? Willst Du ein paar in die Schnauze haben?‘ ein ‚Wortgefecht‘ aus. „Um die vier, die immer näher kamen, zu erschrecken, zog der Angeklagte die von ihm schon während des gesamten Tages mitgeführte Pistole aus der Jackentasche und richtete sie auf die Zeugen. Der Angeklagte wollte sich auf diese Art und Weise wichtig tun. Er ging davon aus, dass er die vier Zeugen mit der Pistole einschüchtern und zur Umkehr bewegen könnte, und forderte sie auf, ihn in Ruhe zu lassen“ (UA Bl. 10). Dabei wollte er weder schießen noch provozieren. Die Jugendlichen wichen zunächst zurück; zwei von ihnen liefen hinter einen vor dem Hauseingang auf der Straße abgestellten Pkw der Marke VW-Golf. Letztlich glaubten sie jedoch, der Angeklagte wolle sie nur mit einer Gaspistole provozieren. Einer rief ‚der schießt doch sowieso nicht‘. Der Aufforderung, ‚kommt, wir nehmen uns die Steine‘, kamen alle vier nach und nahmen aus einem neben dem Hauseingang befindlichen Dränagebeet faustgroße, bis zu 600 g schwere Kieselsteine. Ein Jugendlicher ‚näherte sich sodann mit einem Stein in der Hand erneut dem Angeklagten und rief ihm zu: 'Schieß doch!' Als er das Podest, auf dem der Angeklagte stand, erreicht hatte, stieß dieser ihn dort

---

23 Häder/Klein, ZUMA-Nachrichten 50, 108; Fall q5.

24 Veröffentlicht auch in JZ 1986, 768-768; NJW 1986, 2716-2718.

25 Weshalb Häder/Klein anstelle der neutralen Formulierung „beobachten“ die emotional gefärbte Formulierung „mustern“ verwenden (siehe vorhergehender Absatz), ist nicht erkennbar.

wieder herunter. Nunmehr begannen alle Zeugen, mit den Steinen nach dem Angeklagten zu werfen, der in etwa vier Meter Entfernung mit dem Rücken zur Hauswand im Hauseingang stand und von den Jugendlichen nicht verfehlt werden konnte. Ein Stein traf den Angeklagten am Bein, ein zweiter am Rücken. Einem weiteren Stein, der auf seinen Kopf zuflog, konnte der Angeklagte dadurch, dass er sich wegdrehte und bückte, ausweichen. Diese Reaktion der vier Jugendlichen hatte der Angeklagte nicht erwartet; bereits die Bemerkung 'der schießt doch sowieso nicht', die einer der Jugendlichen abgegeben hatte, bevor alle die Steine aufnahmen, hatte ihn völlig überrascht. Keineswegs hatte er jedoch damit gerechnet, dass die vier ihn trotz der vorgehaltenen Pistole tätlich angreifen würden. Durch den einsetzenden Steinhagel geriet der Angeklagte in Panik. Als ein Stein nur knapp seinen Kopf verfehlte, sah er deshalb‘ - in zutreffender Beurteilung seiner Lage (UA Bl. 23 f) – ‚keinen anderen Ausweg mehr, als die Pistole zu entsichern, den Hahn vorzuspannen und abzudrücken. Der Gefährlichkeit seines Handelns war der Angeklagte sich in diesem Augenblick nicht bewusst; insbesondere dachte er nicht an die möglichen Folgen seines Tuns‘ (UA Bl. 11/12). Der Schuss traf einen der Jugendlichen (den Nebenkläger) in den Wirbelkanal und führte zu einer Querschnittslähmung.“ Der BGH kommt zu dem Ergebnis: „Die Ausführungen, mit denen die Strafkammer den Rechtfertigungsgrund der Notwehr als gegeben erachtet hat, lassen keinen Rechtsfehler erkennen.“

Die zahlreichen Details, die der BGH referiert, entspringen nicht richterlicher Erzählfreude, sondern sind für die juristische Bewertung tragend. Dies gilt z.B. für die Beurteilung einer möglichen Notwehrprovokation durch den Angeklagten und für die Frage der erforderlichen Abstufung der Notwehrmittel nach der Intensität des Angriffs. Es geht hier also nicht um Merkmale, die dem Fall eine emotionale Färbung geben und das Urteil der Befragten hinsichtlich der Forschungsfrage verfälschen könnten, sondern um Kriterien, ohne die eine Bewertung der Entscheidung der Gerichte nicht möglich ist. Auch weitere Fälle dieser Untersuchung sind hinsichtlich der rechtlichen Bewertungskriterien unterdeterminiert und damit sozusagen nicht „entscheidungsreif“.[26] Fall q5 wurde hier deshalb ausführlich dargestellt, weil er von Häder/Klein als besonders „scharfkantig“ eingestuft

26 Z.B. wird der Angreifer in den Fällen q6, q7, q14, q17, q18 und q19 durch Schüsse oder Messerstiche getötet. Dabei enthalten die knappen Fallschilderungen nur teilweise Informationen zu den relevanten Kriterien, ob der Waffengebrauch angedroht wurde bzw. ob eine Drohung möglich und erfolgversprechend gewesen wäre, ob der Angreifer um die Bewaffnung des Notwehr-Übenden wusste, ob der Waffeneinsatz das einzige erfolgversprechende Abwehrmittel war, ob der Angegriffene anstatt der tödlichen Verletzung den Angriff durch eine weniger gravierende Verletzung hätte abwehren können oder ob die Tötung gar absichtlich erfolgte. Zur rechtlichen Relevanz dieser Kriterien vgl. Schönke/Schröder/Lenckner/Perron a.a.O. § 32 Rn. 36, 37 m.w.N.

wird und das Ergebnis, dass nur 8 % der Befragten den Schuss des Angeklagten für gerechtfertigt halten, deshalb als schlagender Beleg für die Ablehnung der Scharfkantigkeit des deutschen Rechts durch die Bevölkerung erscheint.[27] Es dürfte aber unstreitig sein, dass der Anteil derjenigen Befragten, die dem Angeklagten ebenso wie der BGH Notwehr zugebilligt hätten, bei Vorlage des Originalsachverhalts erheblich höher gewesen wäre. Die Antworten zu Fall q5 sind jedenfalls keine valide Messung der Ablehnung der Rechtsprechung des BGH durch die Befragten.

Die Akzeptanz des Rechtsbewährungsprinzips in der Bevölkerung sollte – wie bereits oben erwähnt – speziell mit Fall q19 untersucht werden: Ein Mann, der von fünf jungen Männern zur Herausgabe seiner Geldbörse gedrängt wurde, hatte zu seiner Verteidigung einen Angreifer erschossen.[28] Dieser Sachverhalt beruht in seinen Grundzügen auf dem U-Bahnfall Goetz, der sich 1984 in New York ereignete.[29] Nach deutschem Recht taugt der Fall q19 als Lehrbuchbeispiel für ein unerträgliches Missverhältnis zwischen dem durch die Notwehr geretteten und dem vernichteten Rechtsgut, das die Berechtigung zur Notwehr mangels Gebotenheit entfallen lässt.[30] Es verwundert sehr, dass diese für das Forschungsziel zentrale Frage mit einem Fall erhoben wurde, den deutsche Gerichte nicht entschieden haben und bei dem nach hiesiger Dogmatik weder der Individualrechtsschutz noch das Rechtsbewährungsinteresse eine gegen das Leben des Angreifers gerichtete Verteidigung rechtfertigen.[31] Andererseits zeigen gerade die kontroversen Diskussionen in der Fachwelt, den Medien und der breiten Öffentlichkeit, die der Fall Goetz in den

---

27 Häder/Klein, ZUMA-Nachrichten 50, 102, sowie Tabelle 6 a.a.O. S. 103.

28 Häder/Klein, ZUMA-Nachrichten 50, 103.

29 Tatsächlich bat der Jugendliche Canty den 37-jährigen Goetz nur um fünf Dollar. Goetz schoss auf Canty und die drei schwarzen Jugendlichen in seiner Begleitung. Ein Jugendlicher erlitte eine Querschnittlähmung, getötet wurde keiner; vgl. Fletcher, Notwehr als Verbrechen. Der U-Bahn-Fall Goetz, 1993, 10 f.

30 Vgl. Schönke/Schröder/Lenckner/Perron a.a.O. § 32 Rn. 50. Freilich kommt es auch hier auf die näheren Umstände an. Es können sich z.B. wichtige Dokumente, Medikamente oder hohe Geldsummen in der Börse befinden, relevant sind auch die persönlichen Verhältnisse des Angegriffenen, ob er z.B. Bettler oder Millionär ist; näher hierzu Beulke Jura 1988, 645 f. Beispiele aus der Rechtsprechung: Tödlicher Schuss auf den mit einer Sirupflasche entfliehenden Dieb (OLG Stuttgart DRZ 1949, 42), tödlicher Schuss auf flüchtenden Autoaufbrecher in der Annahme eines Diebstahls von Autoradio, Fahrzeugpapieren und ADAC-Schutzbrief (LG München NJW 1988, 1860, weitere Nachweise bei Schönke/Schröder/Lenckner/Perron a.a.O. § 32 Rn 51.
Ist der Angegriffene nicht durch Notwehr gerechtfertigt, könnte er nach § 33 StGB entschuldigt sein, weil er sich aus Verwirrung, Furcht oder Schrecken zu intensiv verteidigt hat. Zur Entscheidung dieser Frage enthält der Sachverhalt q19 allerdings nicht genügend Anhaltspunkte.

31 Vgl. hierzu Schönke/Schröder/Lenckner/Perron a.a.O. § 32 Rn 50 am Anfang. Zu Unrecht meinen Häder/Klein, ZUMA-Nachrichten 50, 103 a.E., dass das Erschießen des Angreifers in Fall q19 rechtskonform wäre.

USA und darüber hinaus angestoßen hat, dass man den Prinzipien und Grenzen der Notwehr kaum durch spontane Statements im Rahmen einer telefonischen Befragung auf die Spur kommen kann und auf diese Weise auch juristischen Laien nicht gerecht wird.[32]

Häder/Klein verwenden den Begriff Notwehr nicht, weil sie befürchten, dass er von den Befragten unterschiedlich interpretiert wird. Sie haben jedoch keine Bedenken, bei den Fällen q13a und q16 den Befragten den Rechtsbegriff „erforderlich" ohne weitere Erläuterungen zu präsentieren.[33] Zu Fall 16 wird die Frage gestellt: Glauben Sie, dass sich nach dem Gesetz der Angegriffene mit allen erforderlichen Mitteln wehren darf, selbst wenn der Angreifer dabei lebensgefährlich verletzt werden kann, oder dass der Angegriffene sich nur vorsichtig wehren darf, damit der Angreifer möglichst nicht schwer verletzt wird.[34] Nach einer juristischen Standarddefinition ist diejenige Verteidigung erforderlich, die einerseits zur Abwehr des Angriffs geeignet ist und andererseits das nach Art und Maß relativ mildeste Mittel darstellt.[35] Demnach sind im Fall q16 beide Antwortalternativen richtig, da der Notwehr-Übenden den Angriff stets nur auf möglichst schonende und vorsichtige Weise abwehren darf, im Extremfall, wenn mildere Mittel nicht zur Verfügung stehen, aber auch zu lebensgefährdenden Mitteln greifen kann. Die Antwortalternativen bei Frage q16 verkennen anscheinend selbst die Bedeutung des Kriteriums „Erforderlichkeit".

Es bestehen demnach in mehrfacher Hinsicht Zweifel, ob mit den verwendeten Fällen und Fragen die Akzeptanz der Rechtsprechung zum Notwehrrecht und zum Rechtsbewährungsprinzip valide erhoben wurde. Die Kritik betraf einzelne Aspekte der verwendeten Fälle und Fragen, die Konstruktion der Fälle begegnet aber auch grundsätzlichen Bedenken.

Häder/Klein heben die Anwendung der Facettentheorie bei der Bildung der Fälle hervor.[36] Drei Facetten wurden systematisch variiert: das angegriffene Gut, die Stärke der angreifenden Person und der Notwehr-Schaden (leichte Verletzung, schwere Verletzung oder Tod des Angreifers). Bedenken bestehen hinsichtlich der dritten Facette, was sich an Fall q11 exemplarisch zeigen lässt. Ein Wirt wird in seinem Lokal von einer Gruppe Jugendlicher überfallen. Er fällt rücklings und wehrt sich schließlich mit Hilfe einer Eisenstange. Das Verhalten des Wirtes ist nach dem Gesetz gerechtfertigt, wenn er in der

---

32 Ausführlich erörtert wird diese Diskussion in dem brillanten Buch von Fletcher a.a.O. passim.

33 Häder/Klein, ZUMA-Nachrichten 50, 109 f.; auf S. 95 führen Häder/Klein allerdings aus, dass es ratsam erschien, neben der Notwehr auch spezifisch juristische Ausdrücke und Formulierungen wie beispielsweise „Erforderlichkeit" und „Gebotenheit" umzuformulieren.

34 Häder/Klein, ZUMA-Nachrichten 50, 110. Die Frage wurde geringfügig umformuliert, damit sie ohne den Kontext von Fall q16 verständlich ist.

35 Schönke/Schröder/Lenckner/Perron a.a.O. § 32 Rn 34, 36.

36 Häder/Klein, ZUMA-Nachrichten 50, 91 f., insbesondere Abb. 1 auf S. 92.

konkreten Situation über kein milderes wirksames Abwehrmittel verfügte. Das Risiko erheblicher Verletzungen tragen dann die Angreifer, die in Kenntnis dieses Abwehrmittels ihren rechtswidrigen Angriff fortsetzen.[37] Käme es für die Rechtmäßigkeit des Verhaltens auf spätere Verletzungsfolgen an, würde der Angegriffene nicht wissen, ob er sich in zulässiger Weise verteidigt. Denn schon durch ein leichtes Zurückschubsen des Angreifers, kann dieser unglücklich stürzen und sich schwer verletzen (so z.B. in Fall q12). Ob der Angegriffene sich rechtmäßig verhalten hat, könnte erst ex post festgestellt werden, wenn das Geschehen seinen Abschluss gefunden hat. Es ist m.E. rechtlich und ethisch fragwürdig, den Angegriffenen mit diesem Risiko zu belasten und den Angreifer, der sich rechtswidrig verhält, davon freizustellen. Die Vorgaben nach Facette 3 zeigen, dass Häder/Klein diese Auffassung nicht teilen. Die Konstruktion der Fälle nötigte die Befragten, die späteren Folgen einer Verteidigung zu berücksichtigen, ohne dass sie diesen Standpunkt überdenken konnten. Diese „Überrumpelung" der Befragten ist methodisch höchst problematisch. Man hätten ihnen m.E. durch entsprechende Fragen Gelegenheit geben müssen, sich explizit zu dieser Problematik zu äußern. In der Untersuchung wird nicht gemessen, welche Handlungsoptionen man einem in Bedrängnis gebrachten Menschen einräumen soll. Gemessen wird vielmehr, ob man hinterher, wenn alle klüger sind, das Ergebnis insgesamt als gut befindet. Die erhobenen Befunde enthalten deshalb kein Votum der Befragten für die Begrenzung der Notwehr durch eine Güterabwägung.[38]

In der bisherigen Diskussion decken sich ethische und rechtliche Bewertungen weitgehend. Die Untersuchung gilt aber den Grenzen des Notwehr-Rechtes. Es ist deshalb die Frage zu behandeln, wodurch sich das Notwehrrecht von einer Verhaltensethik unterscheidet. Ein wesentliches Kriterium sind die Rechtsfolgen, die an die Entscheidung von Rechtsfragen geknüpft sind, und die ihrerseits zwangsweise durchgesetzt werden können und z.T. sogar müssen. Die Rechtsfolgen der Notwehr betreffen zunächst die Frage, ob der Notwehr-Übende bestraft wird. Darüber hinaus knüpfen sich an die Rechtmäßigkeit der Verteidigung aber weitere Rechtsfolgen. Wer zu Unrecht Notwehr übt, muss für den angerichteten Schaden aufkommen, Schmerzensgeld, Krankenhauskosten, Lohnausfall u.a.m. bezahlen. Bei einem vorsätzlichen Handeln oder bei einem vorsätzlichen Abweichen von gesetzlichen oder behördlichen Vorschriften kommt auch eine bestehende Haftpflichtversicherung nicht für die Schäden auf (§ 4 II Nr. 1, 6 i AHB). Sofern der Angreifer schwere Verletzungen erleidet wie etwa im Fall q5 eine Querschnittlähmung, kann eine unrechtmäßig geübte Notwehr existenzbedrohende Verpflichtungen auslösen. Ohne The-

37 Dies wird üblicherweise dadurch zum Ausdruck gebracht, dass die Erforderlichkeit der Verteidigung nach einem objektiven Urteil ex ante zu bestimmen ist; vgl. Kühl Strafrecht AT, 3. A. 2000, S. 177 Rn 107.

38 so aber Häder/Klein, ZUMA-Nachrichten 50, 102.

matisierung dieser Rechtsfolgen in den Fällen und Antwortvorgaben kann schlechterdings nicht davon gesprochen werden, dass die Befragten in einer Untersuchung zu Rechtsfragen Stellung genommen haben. Die Untersuchung von Häder/Klein beschränkt sich deshalb auf die Erfassung moralischer Wertungen.

Als Fazit bleibt festzuhalten, dass die publizierten Teile der Untersuchung nach der hier vertretenen Auffassung keine validen Aussagen zu der Frage wie weit das Recht das Verhalten regelt und zu der Akzeptanz der Prinzipien der Rechtsprechung in der Bevölkerung zulassen. Der Titel des Aufsatzes ist von den referierten Befunden nicht gedeckt.

Korrespondenzadresse

PD Dr. Arthur Hartmann
Institut für Kriminologie
Sand 6/7
72076 Tübingen

ANMERKUNGEN ZU ARTHUR HARTMANN: VON DER SCHWIERIGKEIT, RECHTSFRAGEN SOZIALWISSENSCHAFTLICH ZU UNTERSUCHEN. REPLIK ZU HÄDER/KLEIN

*MICHAEL HÄDER & SABINE KLEIN**

Auf die Replik von Hartmann wird in vier Schwerpunkten eingegangen. Dabei stehen, dem Charakter der ZUMA-Nachrichten folgend, zunächst vor allem die angesprochenen methodischen Aspekte im Mittelpunkt. Danach wird aus der Sicht der Rechtsdogmatiker auf die in der Replik enthaltenen Bemerkungen reagiert.

Generell ist die Frage, ob bei einem Forschungsanliegen die geeigneten Methoden zum Einsatz gekommen sind, in Abhängigkeit vom Untersuchungsziel zu beantworten. Die Untersuchungsziele der Dresdner Notwehrstudie 2001 wurden in einem facettentheoretischen Ansatz expliziert. Hierauf ist zuerst einzugehen.

Die Prüfung der Validität eines neu entwickelten Instruments hat – jeweils vor dem Hintergrund der angestrebten Untersuchungsziele – mithilfe von Pretests zu erfolgen. Im Rahmen der Multi-Method-Pretestserie (Zwei-Phasen-Pretesting) zum Dresdner Notwehrprojekt kamen vor allem drei kognitive Techniken sowie das sogenannte Standardverfahren zum Einsatz. Letzteres wurde in verschiedenen Phasen der Fragebogenentwicklung wiederholt.

Schließlich wird in der Replik die fundamentale Frage gestellt, ob der gewählte methodische Zugang nicht suboptimal sei, beziehungsweise, ob nicht ein anderer Erhebungsmodus geeigneter gewesen wäre.

---

* Kapitel 4 wurde von den Projektmitarbeitern Knut Amelung und Ines Kilian verfaßt.

Wir behaupten und möchten hier zeigen, dass

1. der facettentheoretische Ansatz und die auf dieser Grundlage vorgenommene Systematisierung der Fälle eine klare Beschreibung der Vorgehensweise und damit der zu erwartenden Ergebnisse erlauben.
2. die umfangreichen Pretestergebnisse belegen, dass das für die Erhebung entwickelte Instrument ausreichend valide ist.
3. mithilfe des im Rahmen des Projekts entwickelten Instruments für eine telefonische Befragung eine optimale Bearbeitung der gewählten Thematik erfolgen kann.

1. Die Festlegung der Vorgehensweise der Dresdner Notwehrstudie durch die Facettentheorie

In der Replik wird angemahnt, eine zu starke Verkürzung der Fälle vorgenommen zu haben. Weiter wird für die Erforschung von Geschlechtsrollen vorgeschlagen, die Fälle weiter zu differenzieren, beispielsweise das Geschlecht der notwehrübenden Person und das Geschlecht der angreifenden Person systematisch zu variieren.

Die Facettentheorie wurde genutzt, um die Ausarbeitung des Erhebungsstandards systematisch zu steuern. Außerdem leistete sie wertvolle Hilfe bei der Ausarbeitung der Pretteststrategie (vgl. unten 2.). Danach liegt der Focus der Studie eindeutig auf (1.) dem jeweils angegriffenen Gut, (2.) auf dem bei der Verteidigung beschädigten Gut und (3.) auf dem Grad der Überlegenheit des Angreifers. Weitere Aspekte bleiben ausdrücklich – so interessant sie auch sein mögen – aus der Betrachtung ausgespart. Das Verständnis dieser Explikation scheint das Kardinalproblem von Hartmann zu sein.

Wie gezeigt (Häder/Klein 2002) war zur Umsetzung dieses Konzepts die Formulierung von 18 Fällen erforderlich, die dann den Befragten zur Entscheidung vorgelegt wurden. Die Aufnahme zusätzlicher Facetten, etwa – wie vorgeschlagen – die Variation des Geschlechts der Angreifer und der angegriffenen Person würde die erforderliche Zahl der zu bewertenden Fälle deutlich erhöhen. Es mag durchaus interessant sein, auch noch das Vorliegen einer möglichen Notwehrprovokation oder eine Abstufung der Notwehrmittel nach der Intensität des Angriffs bei der Erhebung zu berücksichtigen, wie dies von Hartmann ebenfalls vorgeschlagen wird. Eine solche Ergänzung hätte für die Entwicklung des Erhebungsstandards jedoch weitreichende Konsequenzen:

Zusätzlich zu den derzeit berücksichtigten drei Facetten wäre eine Facette „Geschlecht des Angreifers“ (mit zwei Ausprägungen) eine Facette „Geschlecht des Notwehrübenden“ (mit zwei Ausprägungen), eine Facette „Vorliegen von Notwehrprovokation“ (mit mindestens zwei Ausprägungen) und eine Facette „Intensität des Angriffs“ (mit mindestens

zwei Ausprägungen) in den Abbildsatz aufzunehmen. Aus den derzeit (3 x 3 x 2 =) 18 Items umfassenden Fragenkatalog würden dann eine Erweiterung auf mindestens (3 x 3 x 2 x 2 x 2 x 2 =) 144 Items erfolgen müssen. Diese Fälle müssten dann so formuliert werden, dass nicht mehr nur – wie bisher – drei, sondern jetzt sechs Dimensionen von den Befragten kognitiv verstanden und entsprechend bewertet werden können. Ein solches Vorhaben erscheint nach unserer Ansicht die Möglichkeiten der Umfrageforschung weit zu übersteigen (vgl. auch Punkt 3 unserer Bemerkungen).

Der aus der Facettentheorie entlehnte Abbildungssatz stellt ein ausgezeichnetes Mittel dar, um die in einem Fall jeweils enthaltenen Dimensionen zu verdeutlichen und um damit den Anspruch der Erhebung zu demonstrieren. So wird im Rahmen der Studie gewährleistet, dass genau diese Dimensionen von den Befragten bewertet werden. Aussagen, die über die Bewertung der genannten Dimensionen hinausgehen, sind im Rahmen des Projekts nicht beabsichtigt. Auch in anderen Zusammenhängen hat sich ein solches Vorgehen bereits bewährt (vgl. Häder 2000).

Die Fälle sollten, so das Grundanliegen des Projekts, einem repräsentativen Bevölkerungsquerschnitt vorgelegt und von diesem bewertet werden. So waren beispielsweise auch wesentlich ältere Personen und Personen ohne Schulabschluss in unserer Stichprobe vertreten. Wir bezweifeln, ob dieser Personenkreis dazu in der Lage wäre, die sechs Dimensionen, welche dann – wie von Hartmann vorgeschlagen – in den 144 Fällen enthalten sein müssten, valide zu bewerten. Wir können jedoch aufgrund der Pretets (vgl. Punkt 2) belegen, dass unsere drei Dimensionen einschließlich deren Unterteilung von den Befragten reproduziert werden konnten.

In den meisten Fällen, die den Zielpersonen zur Bewertung vorgelegt wurden, agieren Männer. Dies wird von Hartmann bemängelt. Angemerkt werden soll dazu, dass Aktenanalysen bei den Generalstaatsanwaltschaften in verschiedenen Bundesländern (vgl. Kilian/Amelung 2001) ergeben haben, dass sich fast ausschließlich männliche Personen auf das Notwehrrecht berufen. Zu dem gleichen Resultat kommt auch die Auswertung der in der Dresdner Notwehrbefragung enthaltenen Indikatoren, bei der auch die Erfahrungen der Zielpersonen mit Notwehr und mit Notwehrhilfe ermittelt wurden. Von daher entspricht die von uns getroffene Auswahl der zu bewertenden Fälle in etwa der, die in der Realität vermutet werden kann.

Die 18 in die Befragung aufgenommenen Fälle dienten der Überprüfung von Hypothesen, die der Scharfkantigkeit des deutschen Notwehrrechts und der Güterabwägung bei Notwehr gewidmet waren. Entsprechend wurden – nahezu ausschließlich – problematische beziehungsweise strittige Fälle ausgewählt, die dann von den Befragten beurteilt wurden. Diese Fälle korrespondieren mit der Realität insofern, dass sie zumeist auf „wahren“

Fällen beruhen. Sie unterscheiden sich jedoch von der Alltagsrealität in ihrer Drastik und in der Seltenheit ihres Auftretens.

Erwartungen, im Rahmen einer Bevölkerungsbefragung Gerichtsprozesse mit ihrer gesamten Eigendynamik adäquat nachvollziehen zu können, müssen enttäuscht werden. Ein ausführliches Aktenstudium würde beispielsweise selbst mithilfe einer schriftlichen Befragung nicht äquivalent nachgestellt werden können.

2. Multi-Method Pretests als Hilfsmittel zur Validierung der Instrumente

In der Replik wird behauptet, die benutzten Instrumente seien nicht valide. Wir gehen jedoch, gestützt auf eine differenzierte Pretetststrategie, davon aus, dass die drei im Abbildungssatz enthaltenen Facetten einschließlich deren Strukturierung von den Befragten adäquat verstanden wurden.

Unserem Verständnis des Befragtenverhaltens liegt das kognitionspsychologische Modell von Schwarz und Strack (1988) zugrunde. Danach müssen Befragte die ihnen präsentierten Fragen zunächst verstehen, sodann ist zu entscheiden, ob aus dem Gedächtnis bereits eine entsprechende Antwort abgerufen werden kann. Ist dies nicht der Fall, so erfolgt eine Informationsrecherche, welche abgebrochen wird, sobald die befragte Person meint, ein Urteil bilden zu können. Nach der Urteilsbildung erfolgt schließlich das Einpassen des Urteils in das Antwortformat sowie eventuell ein Editieren der Antworten, beispielsweise in Richtung auf eine sozial erwünschte Antwort. Mit den kognitiven Pretestrategien verfügt die Sozialforschung über Instrumente, um diesen Antwortprozess transparent zu machen. So wird es möglich zu prüfen, ob die Zielpersonen die ihnen vorgelegten Fragen tatsächlich in dem Sinne interpretieren und beantworten, der von den Forschern beabsichtigt wird.

Für die ersten Pretests wurden kognitive Techniken entwickelt. Diese wurden in einen Evaluationsfragebogen eingearbeitet, der die Anwendung der kognitiven Techniken weitgehend standardisiert. Dies erfolgte vor allem, um bei dem Test auch erfahrene Pretestinterviewer einsetzen zu können, welche die Projektbearbeiter bei der Überprüfung ihres Instrumentenentwurfs unterstützen. Wären die kognitiven Techniken innerhalb eines offenen Tiefeninterviews angewandt worden, hätten nur die mit dem Projekt vertrauten Experten diese Interviews führen können. Zudem hätte sich die Auswertung schwieriger und langwieriger gestaltet (vgl. Prüfer/Rexroth 2000).

Innerhalb des Pretests fanden drei unterschiedliche kognitive Methoden Anwendung. Dabei handelt es sich um die Labormethoden ‚Paraphrasing', um ‚Special Comprehension

Probes‘ sowie um ‚General Probes‘. Alle Interviewer wurden in der technischen Handhabung des Pretestinstrumentes unterwiesen, jedoch nicht inhaltlich beeinflusst. Die Neutralität des Interviewers gegenüber dem Befragten und den Befragungsinhalten ist neben der Standardisierung das wichtigste Kriterium zur Herstellung gleicher Bedingungen und unabdingbar zur Gewährleistung der Vergleichbarkeit der Antworten, und damit der Resultate (vgl. Prüfer/Stiegler 2002).

Der Auszug aus dem Evaluationsfragebogen (vgl. Häder/Klein 2002) illustriert dies. Es ging darum, projektfremde Interviewer mit einem Pretestinstrument auszustatten, um ihnen eine standardisierte Informationsbeschaffung zu ermöglichen. Die Vermittlung von Vorurteilen gegenüber dem deutschen Notwehrrecht, wie Hartmann anzunehmen scheint, war nicht unsere Absicht.

Außerdem wurden in verschiedenen Phasen der Fragebogenentwicklung auch verschiedene Standardpretets (als telefonische Befragung) veranstaltet.

Die Auswertung dieser Tests beziehungsweise die aus den einzelnen Testergebnissen für die Überarbeitung des Instruments vorgenommenen Veränderungen begründen im Weiteren unsere Auffassung, ein unter den gegebenen Umständen optimales Instrument entwickelt zu haben, um die oben (vgl. 1) genannten Sachverhalte abbilden zu können.

An dieser Stelle soll nochmals hervorgehoben werden: Die Zielpersonen unserer Voruntersuchungen, unter ihnen auch ältere Menschen sowie Personen ohne Schulabschluss, vermochten es in den Pretests (zum Beispiel beim ‚Paraphrasing‘), die in den Fällen enthaltenen Dimensionen aus dem Gedächtnis zu repetieren. Einige Fälle mussten allerdings im Ergebnis der Voruntersuchungen umformuliert werden, da sie sich als nicht eindeutig oder zu schwierig erwiesen.

Die am Pretest Beteiligten waren darüber hinaus dazu in der Lage, eine Reihe an ergänzenden Nachfragen (‚Probings‘) adäquat zum Untersuchungskonzept zu beantworten (zu den Details vgl. Häder/Klein 2002).

3. Telefoninterviews oder postalische Befragungen?

In der Replik wird empfohlen, besser einen anderen Zugang zum Untersuchungsfeld zu wählen, beispielsweise eine schriftliche Befragung. Dieser Vorschlag basiert unter anderem auf dem vermuteten Auftreten von Interviewereffekten im Rahmen von Telefonbefragungen.

Diese Replik veranlasst dazu, zunächst allgemein über die Möglichkeiten und Grenzen der Umfrageforschung zu reflektieren.

Das Dresdner Notwehrprojekt bewegt sich an der Nahtstelle zwischen zwei Wissenschaftsdisziplinen, der Umfrageforschung und der Rechtswissenschaft. Gerade bei solchen Projekten ist es wichtig, den Entwicklungsstand der jeweils anderen Seite zur Kenntnis zu nehmen. Nur so können beispielsweise sowohl unrealistisch weitgesteckte Erwartungen an die Ergebnisse von Umfragen aber auch das Übersehen der inzwischen immens angewachsenen Möglichkeiten der Umfrageforschung vermieden werden. Beiden Tendenzen unterliegt jedoch Hartmann in seiner Replik.

So würde es sich *erstens* um eine unrealistische Erwartung an die quantitative Sozialforschung handeln, dass sie (völlig) objektive Daten produzieren könne, die am besten ohne jeden subjektiven Einfluss erzeugt werden. Wenn man akzeptiert, dass es in der Sozialforschung einen Bedarf an quantitativen Daten über die Verteilung bestimmter Ausprägungen in einer Grundgesamtheit gibt, so muss auch akzeptiert werden, dass die Sammlung solcher Daten nicht ohne äußere Einflüsse erfolgen kann. Strenggenommen kann man sich – dies haben nicht zuletzt Test-Retest-Studien am ZUMA (vgl. Porst/Zeifang 1987, Zeifang 1987a und 1987b) gezeigt – nicht einmal sicher sein, ob die Angabe eines bestimmten Berufes im Datensatz tatsächlich mit der Wirklichkeit übereinstimmt. Nun gibt es zwei Wege mit diesem Dilemma umzugehen: Entweder ganz auf die quantifizierende Sozialforschung zu verzichten oder mögliche Fehlerquellen bei der Interpretation von Umfragedaten zu beachten und zugleich schrittweise die zur Verfügung stehenden Instrumente zu verfeinern. Dann gilt es, das jeweils optimale Design für ein bestimmtes Untersuchungsanliegen zu finden und sich über die dabei eingegangenen Kompromisse Klarheit zu verschaffen. Letzteres geschah in unserem Notwehrprojekt.

Die Umfrageforschung ist *zweitens* vor allem in der letzten Zeit zum Gegenstand von intensiven Forschungsbemühungen geworden. Verwiesen sei auf Schwerpunkte wie beispielsweise das Non-Response-Problem (vgl. Dillman 2000, Groves, u.a. 2001), auf die Stichprobenziehung (vgl. Gabler/Häder 2002), auf die Pretestsdesigns (vgl. Prüfer/Rexroth 2000), auf die kognitionspsychologischen Bemühungen, um die Begründung von Fragereihenfolgeeffekten und um den Einfluss von Antwortskalen auf das Antwortverhalten herauszuarbeiten (vgl. Maynard u.a. 2002; Tourangeau u.a. 2000; Sudman u.a. 1996; Schwarz 1996) und schließlich auf Bemühungen, die der internationalen Vergleichbarkeit von Umfrageergebnissen gewidmet sind (vgl. Harkness u.a. 2002). Es ist versucht worden, dieses Wissen bei der Konzipierung unserer Umfrage zu berücksichtigen.

Die Dresdner Notwehrstudie hat sich einer Telefonbefragung in einem CATI-Labor bedient. Die in der Replik vorgeschlagene schriftliche – gemeint ist wohl eine postalische – Befragung, stellt dagegen, sieht man einmal von Internet-Befragungen ab, die denkbar schlechteste Lösung dar. Erinnert sei an die fehlenden Möglichkeiten zur Stichprobenzie-

hung für eine postalische allgemeine Bevölkerungsbefragung in Deutschland, an die fehlenden Möglichkeiten zu kontrollieren, wer letztendlich den Fragebogen ausfüllt sowie an die hohe Unsicherheit bei der Bestimmung der zu erwartenden Response-Rate beziehungsweise des zu erwartenden Rücklaufes. Hier hätte bereits jedes einzelne Argument ausgereicht, um auf eine postalische Befragung verzichten zu müssen. Dazu kommen aber auch noch weitere Argumente wie die fehlenden Möglichkeiten für Rückfragen, die notwendigerweise simple Gestaltung des Fragebogens, welche dazu zwingt, weitgehend auf Filter zu verzichten und so weiter.

Auch das Argument, bei postalischen Befragungen könne man jeden Interviewereinfluss ausschließen, ist nicht sonderlich stark, insbesondere ist nicht bekannt, welchen Einfluss anwesende dritte Personen beim Ausfüllen des Fragebogens ausüben, in welchem Kontext also der Fragebogen ausgefüllt wird, etwa zuhause alleine mit der Absicht überlegt und ernsthaft zu antworten oder vielleicht gar in geselliger Runde. „Selbst eine Aussage wie: ‚Den Fragebogen habe ich mit meinem Sportkameraden nach dem Training ausgefüllt, wir waren durch ein paar Bier schon ziemlich angeheitert und haben furchtbar dabei gelacht' ist weder zu bestätigen noch zu dementieren" (Klein/Porst 2000: 13).

Hartmann hält Telefonbefragungen lediglich für geeignet, wenn es um Vorlieben beim Kauf von Waschmitteln geht. Ein solches Quick-and-Dirty-Image entspricht einem Forschungsstand der empirischen Sozialforschung aus den 80er Jahren. Es ist schon erstaunlich, dass solche Ansichten inzwischen noch nicht grundsätzlich revidiert worden sind. In Publikationen, die in Deutschland zu Beginn der 90er Jahre erschienen sind (vgl. Frey u.a. 1990, Lavrakas 1993) und in den USA bereits aus der Mitte der 70er Jahre stammen (vgl. Rogers 1976), wird dies ausreichend belegt.

Im Jahr 2000 wurden dann auch 45 Prozent aller Interviews, die von den Mitgliedsinstituten des Arbeitskreises Deutscher Markt- und Sozialforschungsinstitute geführt wurden, per Telefon abgewickelt (vgl.: http://www.adm-ev.de). Damit stellt dieser Mode, noch vor den face-to-face Befragungen, den häufigsten Zugang dar.

Für diese Tatsache können verschiedene Argumente verantwortlich gemacht werden, die inzwischen auch Eingang in jedes sozialwissenschaftliche Methodenlehrbuch gefunden haben, die Hartmann jedoch offenbar bisher entgangen sind. Vor allem ist zu verweisen auf einen, gegenüber face-to-face Befragungen geringeren Interviewereinfluss (vgl. Frey u.a. 1990; Porst 2000), auf die sehr guten Möglichkeiten der Interviewerkontrollen durch die Supervisoren, auf die inzwischen zur Verfügung stehenden Strategien zur Ziehung einer Zufallsstichprobe (vgl. Gabler/Häder 2002), auf die Möglichkeit bei Rückfragen den Supervisor zu konsultieren, auf die Möglichkeit, Indikatoren zufällig zu permutieren und so weiter.

Für die Entscheidung, die Dresdner Notwehrstudie als CATI-Befragung zu veranstalten, gab es damit keine Alternative.

Dass bei telefonischen Befragungen ein geringerer Interviewereinfluss zu erwarten ist als bei face-to-face Befragungen, ist unter anderem einer größeren räumlichen Distanz und dem fehlenden visuellen Kontakt zwischen Interviewer und Befragten zuzuschreiben, so liegen beispielsweise keine Informationen über das Alter, die äußerliche Erscheinung und die Kleidung des Interviewers vor. Dennoch stellt sich die Frage, welchen Einfluss die Interviewer in der Dresdner Notwehrbefragung gehabt haben können. Prinzipiell kann ein Einfluss der Interviewer auch bei Telefonbefragungen nicht ausgeschlossen werden. Deshalb muss es darum gehen, erstens diesen Einfluss so gering wie möglich zu halten und zweitens in den gewonnenen Daten dann nach Hinweisen auf einen solchen Einfluss zu suchen.

Bei den Strategien, die im Notwehrprojekt eingesetzt wurden, um den Interviewereinfluss zu minimieren, handelt es sich um spontanes und um regelmäßiges, stichprobenartiges Mithören der Interviews sowohl durch einen Supervisor des beauftragten Erhebungsinstituts[1] als auch durch einzelne Mitglieder der Projektgruppe, um eine heterogene Zusammensetzung des Interviewerstabes sowie um das automatische Einspielen der Fälle von einem Tonband, um jeden Unterschied, etwa in der Betonung bestimmter Worte, zu vermeiden.

Tabelle 1 gibt Auskunft über die Struktur des Interviewerstabes, den das von uns beauftragte Institut, USUMA, bei der Befragung eingesetzt hat.

52 Prozent der Interviews wurden von Frauen geführt. Das Durchschnittsalter der Interviewer beträgt 37 Jahre, wobei der älteste Interviewer 75 Jahre und der jüngste 18 Jahre alt war. Es ist davon auszugehen, dass dieser Personenkreis aufgrund seiner heterogenen Struktur ebenfalls heterogene Ansichten zur untersuchten Problematik vertritt. Mögliche Fehler, so wird weiter angenommen, neutralisieren sich damit weitgehend. Um bei der Datenauswertung trotzdem vorhandene Interviewereinflüsse ermitteln zu können, sind Alter, Geschlecht und Bildungsstand der jeweiligen Interviewer dem Datensatz zugespielt

1 Die Studie wurde von der USUMA GmbH in Berlin nach den ADM-Standards (Quelle: http://www.adm-ev.de/) durchgeführt. Den Veranstaltern ist – mit Ausnahme der bei USUMA für das Projekt zuständigen Projektleiterin, welche in die Diskussion der Pretestergebnisse einbezogen wurde – kein Interviewer persönlich bekannt. Eine Ausnahme bilden lediglich zwei Interviewer (von ca. insgesamt 40 USUMA-Interviewern, die für die Notwehrstudie telefoniert haben), die im Rahmen einer Expertendiskussion ihre beim Pretest mit dem Instrument gemachten Erfahrungen vortrugen. Die Vermutung von Hartmann, die Interviewer seien mit den Positionen der Wissenschaftler vertraut gewesen, ist von daher völlig haltlos.

worden. An dieser Stelle sollen einige Befunde entsprechender Nachforschungen präsentiert werden.

Tabelle1: Höchster Bildungsabschluss der Interviewer pro Interview in der Dresdner Notwehrstudie, „Was trifft auf Sie zu ...“

	Häufigkeit	Prozent
... noch Schüler(in),	108	3,1
... ohne Hauptschulabschluss	3	,1
... Hauptschulabschluss	249	7,2
... Realschulabschluss	1.341	38,7
... Abitur (früher 12. Klasse)	834	24,1
... Fachhochschulreife	29	,8
... abgeschlossenes Fachschulstudium	237	6,8
... abgeschlossenes Hochschulstudium	661	19,1
Gesamt	3.463	100,0

Bei vier (q2, q6a, q8 und q9a) von 18 Fällen wurde ein signifikanter ($p < .05$) Unterschied zwischen den von männlichen und weiblichen Interviewern ermittelten Antworten festgestellt. Dabei ergab sich, dass bei Interviewerinnen jeweils häufiger die Antwortvorgabe „nicht gerechtfertigt“ aufgetreten ist. Bei nur zwei (q1 und q2a) von 18 Fällen wurde ein solcher signifikanter Unterschied zwischen von formal höher und von formal niedriger gebildeten Interviewern ermittelt. Dabei stellte sich heraus, dass höher gebildete Interviewer jeweils häufiger die Antwort „gerechtfertigt“ ermittelten.

Diese Befunde stellen eine wichtige Ergänzung für die – an anderer Stelle noch weiter vorzunehmende – inhaltliche Auswertung der Daten dar. Wir gehen jedoch nicht davon aus, dass die ermittelten Differenzen den gewählten Untersuchungsansatz, die Telefonbefragung, infrage stellen.

4. Einige Argumente aus rechtsdogmatischer Sicht

Unsere Ausführungen, das deutsche Notwehrrecht gehe davon aus, dass das Recht dem Unrecht „niemals“ zu weichen habe und eine „Schlägermoral“ fördere sind – das ist zugestanden – missverständlich, weil zu hart. Die Formel von der „Schlägermoral“ gibt eine seit dem Ende des 19. Jahrhunderts geübte Kritik wieder, die zumindest seit dem zweiten Weltkrieg so nicht mehr berechtigt ist, wird doch das Rechtsbewährungsprinzip von vie-

len auch zur Beschränkung des Notwehrrechts herangezogen, wo die Ausübung dieses Rechts die Bewährungsfunktion nicht mehr erfüllen kann. Deshalb gibt auch die Formel, dass das Recht „niemals" dem Unrecht weichen müsse, nicht den inzwischen erreichten Stand der Notwehrdogmatik exakt wieder. Es ist eben, wie Hartmann auch schreibt, nur ein „Grundsatz", dass nach deutscher Auffassung das Recht dem Unrecht „nicht" zu weichen brauche.

Beim Steinwerfer-Fall (q5) können wir nicht nachvollziehen, inwiefern die Wiedergabe im Fragebogen die vom Bundesgerichtshof beurteilte Fallkonstellation entscheidend verkürzt. Das Provokationsmoment wird in der Schilderung angesprochen, die ansteigende Gefährlichkeit des Angriffs dargetan und der mehr einer Warnung als einer gezielten Verteidigung ähnelnde Charakter der Schüsse angedeutet – alles natürlich ohne Gebrauch rechtstechnischer Begriffe. Die rechtserheblichen Merkmale der Fallschilderung des Bundesgerichtshofs, die nicht in den Fragebogen aufgenommen wurden, die „Panik" des Angegriffenen und das Nichtbedenken der Tatfolgen, betreffen nicht des *Recht* zur Notwehr, sondern die Schuld im Rahmen des Paragraphen 33 des Strafgesetzbuches. Mit der Abgrenzung von Unschuld und Schuld wollten wir aber die Frage nicht belasten.

Ebenso ist der New Yorker U-Bahn Fall gewiss kein „Lehrbuchbeispiel" für eine Notwehreinschränkung (q19) wegen eines unerträglichen Missverhältnisses zwischen Anlass und Folge der Verteidigung. Hartmann übersieht, dass die Angreifer nicht nur ein Vermögensinteresse des Notwehr Übenden bedrohten, sondern auch Persönlichkeitsgüter, und außerdem ist daran zu erinnern, dass selbst das Zurückweichen bei der Bedrohung von Sachgütern in Deutschland keineswegs unumstritten ist.

Ganz neben der Sache liegt die Behauptung Hartmanns, dass in den Fragen q13a und q16 den rechtstechnischen Begriff der „Erforderlichkeit" verkannt wird. Es ist aus der Frage klar zu erkennen, dass dieser Begriff hier nicht rechtstechnisch verwendet wird.

Neben der Sache liegt schließlich auch Hartmanns Behauptung, dass die Autoren der Untersuchung das „Rechtsbewährungsprinzip" ablehnen. Sie sehen in ihm lediglich ein im Wortsinn „*frag*"würdiges, der Untersuchung wertes Merkmal des deutschen Notwehrrechts, dessen Akzeptanz oder Nicht-Akzeptanz durch die Bevölkerung ebenso zur „Ablehnung" dieses Merkmals wie zur Auslösung volkspädagogischer Aufklärung mit dem Ziel erhöhter Anerkennung dieses Prinzip herangezogen werden kann.

5. Fazit

Die Dresdner Notwehrstudie wollte im Rahmen einer Umfrage die Ansichten eines breiten Bevölkerungsquerschnitts zum Thema Notwehr erfragen. Die konkreten Schwerpunkte der Studie wurden in einem facettentheoretischen Ansatz zuvor genau umrissen (vgl. Punkt 1). Eine Konkurrenz zu unter Umständen mehrtätigen Gerichtsverhandlungen, an denen unter anderem eine Reihe geschulter Spezialisten beteiligt ist, war nicht beabsichtigt. Die Schwierigkeit, Rechtsfragen sozialwissenschaftlich zu untersuchen, konnte jedoch durch den benutzten facettentheoretischen Ansatz bewältigt werden.

Mithilfe einer Pretestserie ist erreicht worden, dass die im Untersuchungsansatz fixierten Ziele durch das Erhebungsinstrument valide erhoben werden. Eine telefonische Befragung stellt sich als optimale Lösung dar, um das Instrument einzusetzen. Die dabei möglicherweise auftretenden Interviewereinflüsse müssen bei der Datenauswertung berücksichtigt werden.

Literatur

Dillman, D.A., 2000: Mail and Internet Surveys. The Tailored Design Method. 2. Auflage. New York: Wiley.

Gabler, S./Häder, S., (Hrsg.) 2002: Telefonstichproben: Methodische Innovationen und Anwendungen in Deutschland. Münster: Waxmann.

Groves, R.M./Dillman, D.A./Eltinge, J.L./Little, R.J.A. (Hrsg.) 2001: Survey Nonresponse. New York: Wiley.

Häder, M., 2000: Die Zukunft des Mobilfunks: Facettentheoretisch gestützte Strukturierung einer Delphi-Studie. S. 99-123 in: P. Ph. Mohler/P. Lüttinger (Hrsg.), Querschnitt. Festschrift für Max Kaase. Mannheim: ZUMA.

Häder, M./Klein, S., 2002: Wie wenig das Recht unser Verhalten regelt. ZUMA-Nachrichten 50: 86-113.

Harkness, J./van de Vijer, F./Mohler, P.Ph., (Hrsg.) 2002: Cross-Cultural Survey Methods. New York: Wiley.

Kilian, I./Amelung, K., 2001: Fragenentwicklung aus juristischer Sicht und Aktenanalyse innerhalb des Projekts „Einstellungen über und potentielle Verhaltensintentionen bei Notwehr in der Allgemeinbevölkerung der Bundesrepublik Deutschland“. Zwischenbericht zum Projekt für die VW-Stiftung.

Klein, S./Porst, R., 2000: Mail Surveys. Ein Literaturbericht. ZUMA-Arbeitsbericht 2000/10. Mannheim: ZUMA.

Lavrakas, P.J., 1993: Telephone Survey Methods. Selection, Sampling, and Supervision. 2. Auflage. Newbury Park: Sage.

Maynard, D.W./Houtkoop-Steenstra, H./Schaeffer, N.C./van der Zouven, J. (Hrsg.) 2002: Standardization and Tacit Knowledge. Interaction and Practice in the Survey Interview. New York: Wiley.

Porst, R./Zeifang, K., 1987: Wie stabil sind Umfragedaten? Beschreibung und erste Ergebnisse der Test-Retest-Studie zum ALLBUS 1984. ZUMA-Nachrichten 20: 8-31.

Porst, R., 2000: Praxis der Umfrageforschung. Heidelberg: Teubner.

Prüfer, P./Rexroth, M., 2000: Zwei-Phasen-Pretesting. ZUMA-Arbeitsbericht 2000/08. Mannheim: ZUMA.

Prüfer, P./Stiegler, A., 2002: Die Durchführung standardisierter Interviews. Ein Leitfaden. ZUMA How-to-Reihe, Nr. 11. ZUMA: Mannheim.

Schwarz, N., 1996: Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation. Mahwah: Erlbaum.

Schwarz, N./Strack, F., 1988: The Survey Interview and the Logic of Conversation: Implications for Questionnaire Construction. ZUMA-Arbeitsbericht 1988/03. Mannheim: ZUMA.

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

Tourangeau, R./Rips, L./Rasinski, K.A., 2000: The Psychology of Survey Response. Cambridge University Press: Cambridge.

Zeifang, K., 1987a: Die Test-Retest-Studie zum ALLBUS 1984. Tabellenband. ZUMA-Arbeitsbericht 1987/01. Mannheim: ZUMA.

Zeifang, K., 1987b: Die Test-Retest-Studie zum ALLBUS 1984. Abschlußbericht. ZUMA-Arbeitsbericht 1987/02. Mannheim: ZUMA.

Anhang

Es erfolgt eine Auflistung der in die Befragung und in den Artikel eingeflossenen Fälle samt Quellennachweis, soweit die Fälle auf Quellen beruhen.

Fallnummer	Quellennachweis	Im Artikel zitierte Beispielfälle:
q1	BayObLG NJW 1963/824	Beispielfall 4
q4	Hillenkamp, T. (1990): 30 Probleme aus dem Strafrecht. 6. Auflage. Frankfurt: Metzner: 29	
q5	BGH NstZ 1986, 357	Beispielfall 1
q5a	BGH NstZ 1986, 357 - abgewandelt	
q6	BGH GA 1956, 49	
q6a	BGH GA 1956, 49 - abgewandelt	
q7	BGE NStZ 1983, 117	
q7a	BGE NStZ 1983, 117 - abgewandelt	
q8	BayObLG NJW 1993, 211	Beispielfall 3
q9	BayObLG NJW 1962, 1782	
q9a	BayObLG NJW 1962, 1782 - abgewandelt	
q10	Kunz, K.-L. (1984): Das strafrechtliche Bagatellprinzip. Eine strafrechtsdogmatische und kriminalpolitische Untersuchung. Berlin: Dunker und Humboldt: 539.	
q11	BGH StV 1982, 219	Beispielfall 2
q12	BGHSt 3, 217	
q12a	BGHSt 3, 217 – abgewandelt	
q13	BGH GA 1965, 147	
q14	BGH NJW 1980, 2263	
q15	OLG Frankfurt VRS 40, 424	
q15a, q15b	OLG Frankfurt VRS 40, 424 – abgewandelt	
q16	BGHSt 24, 356ff	
q17, q17b	BGH NJW 1975, 62 – abgewandelt	
q17a	BGH NJW 1975, 62	
q19	In Anlehnung an: Fletcher, G.P. (1993): Notwehr als Verbrechen. Der U-Bahn-Fall Goetz. Frankfurt: Suhrkamp	

ALLGEMEINE BEVÖLKERUNGSUMFRAGE DER SOZIALWISSENSCHAFTEN (ALLBUS): PLANUNG ZUKÜNFTIGER BEFRAGUNGSSCHWERPUNKTE

HANS-JÜRGEN ANDREß

Die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) ist eine wissenschaftliche Dienstleistung der „Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen“ (GESIS) für Forschung und Lehre in den Sozialwissenschaften. Der ALLBUS erhebt seit 1980 regelmäßig im Abstand von zwei Jahren Einstellungen, Verhaltensweisen und Sozialstruktur der Bevölkerung in der Bundesrepublik Deutschland. Das Zentralarchiv für empirische Sozialforschung in Köln stellt die Daten jeder Erhebung unmittelbar im Anschluss an ihre Aufbereitung zur Verfügung.

Das Fragenprogramm jeder ALLBUS-Umfrage besteht aus drei Teilen: ein bis zwei inhaltliche Schwerpunktthemen, Einzelfragen bzw. kürzere Itembatterien zu unterschiedlichen Fragestellungen sowie ausführliche demographische Informationen. Schwerpunktthemen werden in der Regel alle 10 Jahre repliziert, eine Vielzahl von Einzelfragen bzw. kürzeren Itembatterien wird alle zwei bzw. vier Jahre wiederholt.

Auf seiner Sitzung Ende 2001 hat der wissenschaftliche Beirat[1] des ALLBUS das Schwerpunktthema für den ALLBUS 2004 festgelegt. Es lautet „Soziale Ungleichheit“ und soll, gemäß der Konzeption des ALLBUS, die Befragungsschwerpunkte aus den ALLBUS-Erhebungen der Jahre 1984 und 1994 replizieren. Für die konkrete Ausgestaltung des Erhebungsinstruments wurde ein „Call for Proposals“[2] an mehrere einschlägige

---

1 Mitglieder des wissenschaftlichen Beirates sind zur Zeit: Prof. Dr. H.J. Andreß (Bielefeld, Vorsitz), Prof. Dr. W. Bürklin (Potsdam), Prof. Dr. A. Diekmann (Bern), Prof. Dr. H. Feger (Berlin), Prof. Dr. J. Huinink (Rostock), Prof. Dr. H. Meulemann (Köln) und Dr. H. Solga (Berlin). Sie werden von der Mitgliederversammlung des Zentrums für Umfragen, Methoden und Analysen (ZUMA, Mannheim) berufen.

2 http://www.gesis.org/Dauerbeobachtung/Allbus/Aktuelles/callfor.htm (10.7.2002).

Mailing-Listen, verschiedene Sektionen der Deutschen Gesellschaft für Soziologie sowie an ausgewählte Forscher und Forscherinnen versandt. Mit diesem Aufruf sollte der Profession die Gelegenheit gegeben werden, neue Fragen für den ALLBUS vorzuschlagen.

Insgesamt gingen zwölf Vorschläge und mehrere Kommentare ein, die der wissenschaftliche Beirat auf seiner Sitzung im Juli 2002 zu bewerten hatte. Sie wurden als überaus anregend und innovativ wahrgenommen.

Laut Ausschreibungstext konnten die einzureichenden Vorschläge von Einzelfragen bis zu größeren Fragemodulen reichen, wobei eine maximale Befragungszeit von 15 Minuten vorgegeben war. Sie sollten theoriegeleitet sein, es sollte die sozialwissenschaftliche Relevanz der vorgesehenen Konzepte und Dimensionen begründet und die Existenz einschlägiger, für allgemeine Bevölkerungsumfragen tauglicher Indikatoren belegt werden. Wenn für das vorgeschlagene Thema noch keine Indikatoren vorliegen, sollten die neu zu entwickelnden Fragen skizziert werden.

Von der Form her unterschieden sich die eingereichten Vorschläge zum Teil erheblich. Die Bewertung durch den wissenschaftlichen Beirat fiel immer dann schwer, wenn entweder die theoretische Begründung nicht ausgearbeitet war oder wenn die zu verwendenden Fragen nicht benannt wurden. In zukünftigen Ausschreibungsrunden sollte daher eine strengere Vorauswahl getroffen werden.

Inhaltlich gesehen konzentrierten sich die Vorschläge auf den Zusammenhang von Lebensstilen und klassischen Konzepten von Klasse und Schicht, auf neue Formen sozialer Ungleichheit (Stichworte: Digital Divide, kognitive Kompetenz), auf die Entstehung und Perpetuierung sozialer Ungleichheiten auf individueller und familiärer Ebene in Form von Bildungsaspirationen und -strategien sowie auf die Effekte sozialer Ungleichheit in verschiedenen Lebensbereichen (z.B. Gesundheit). Einige Vorschläge plädierten auch für Modifikationen und Ergänzungen der ALLBUS-Demographie, die gerade für die Analyse sozialer Ungleichheiten von vielen Forschern und Forscherinnen genutzt wird.

Der wissenschaftliche Beirat stand vor der schwierigen Aufgabe, aus dieser Vielzahl von innovativen Vorschlägen auszuwählen, gleichzeitig bewährte Items aus den Jahren 1984 und 1994 zu replizieren, alles mit einer umfangreichen Demographie abzuschließen und dabei insgesamt nicht mehr als 50 Minuten Befragungszeit zu verbrauchen. Zunächst wurde beschlossen, die vielen Vorschläge zur Modifikation der ALLBUS-Demographie sehr genau zu prüfen und insgesamt die soziodemographischen Informationen für dieses Schwerpunktthema auszubauen. In einem zweiten Schritt wurde dann über die Vorschläge zu Lebensstilen und Typen der Lebensführung beraten. Da das Lebensstilkonzept bereits im ALLBUS 1998 erhoben wurde, wurde die Strukturentscheidung getroffen, diese Vor-

schläge nicht in die Auswahl einzubeziehen, sondern auf die Replikation der 98er-Erhebung im Jahr 2008 zu verweisen.

Für die verbliebenen Vorschläge fiel die Auswahl nicht leicht. Aufgrund der Bedeutung von Information und Wissen für die moderne Gesellschaft wurde die Hypothese einer digitalen Spaltung der Gesellschaft sowohl unter theoretischen als auch unter empirischen Gesichtspunkten als prioritär angesehen. Dem Thema „Digital Divide" wurden daher zehn Minuten Befragungszeit im ALLBUS 2004 eingeräumt, wobei die konkreten Frageformulierungen noch in Zusammenarbeit mit den vorschlagenden Forschern durch die ALLBUS-Abteilung bei ZUMA ausgearbeitet werden müssen. Die Nutzung traditioneller Medien wie Zeitung, Buch und Fernsehen soll dabei auf jeden Fall auch erfasst werden. Interessant und bedeutsam schien dem Beirat auch der Zusammenhang zwischen Gesundheit und sozialer Ungleichheit. Er möchte jedoch zuvor klären, ob dieses Thema nicht bereits in anderen Bevölkerungsumfragen besser erfasst wird. Zwei weitere Vorschläge – politische Kompetenz als Teilaspekt kognitiver Kompetenz sowie mikrotheoretische Determinanten von Bildungsungleichheiten und Mobilitätsprozessen – erschienen aufgrund ihrer theoretischen Begründungen ebenfalls als untersuchungsbedürftig, litten jedoch unter dem Umstand, dass mögliche Fragen zu dem jeweiligen Themenkomplex nur angedeutet waren und daher in Zusammenarbeit mit den Beiratsmitgliedern und vorschlagenden Forschern durch die ALLBUS-Abteilung erst ausgearbeitet werden müssen. Alle drei Vorschläge (Gesundheit, politische Kompetenz, mikrotheoretische Determinanten) wurden daher auf eine Prüfliste gesetzt, wo sie mit einer ausgebauten Demographie um die verbleibende Befragungszeit konkurrieren müssen. Welches der genannten Themen dabei in welcher Form in das Erhebungsinstrument übernommen wird, wird der Beirat auf seiner Herbstsitzung 2002 auf Basis der ausgearbeiteten Fragen entscheiden.

Aufgrund des positiven Ergebnisses der Ausschreibung wurde beschlossen, das gewählte Verfahren auch bei zukünftigen ALLBUS-Erhebungen einzusetzen, dabei aber den Aufruf noch breiter zu streuen als dies beim jetzigen Mal der Fall war. Der Beirat behält sich allerdings das Recht vor, Themen- und Fragenvorschläge zu modifizieren bzw. Lücken zu benennen, die im Zuge der konkreten Erstellung des Erhebungsinstruments noch ausgearbeitet werden müssen.

Der Beirat dankt allen Kollegen und Kolleginnen, die sich mit Vorschlägen an dem „Call for Proposals" beteiligt haben. Weitere Anregungen zur Planung und Durchführung des ALLBUS sind jederzeit willkommen.

EDITION VON LEBENSVERLAUFSDATEN: ZUR RELEVANZ EINER SYSTEMATISCHEN EINZELFALLBEARBEITUNG BEI STANDARDISIERTEN BEFRAGUNGEN

STEFFEN HILLMERT

1. Einleitung

Ein empirisches Forschungsprojekt lässt sich idealtypisch in Phasen gliedern (etwa Schnell et al. 1995: 7-13), die von der Problemstellung und Konzeptionalisierung über die Auswahl der Untersuchungseinheiten, die Erhebung, Erfassung und Analyse der Daten bis zur theoretischen Reflexion und Darstellung der Forschungsergebnisse reichen. Auf jeder dieser Stufen können Fehler auftreten, die die Gültigkeit der Ergebnisse einschränken, und es sind jeweils spezifische Vorkehrungen dagegen möglich. Ziel sollte es aber bleiben, den *Gesamtfehler* möglichst klein zu halten. Während in der Methodologie der Forschung großes Gewicht auf die Datenanalyse und ihre Methoden gelegt wird, erhalten vorgelagerte Stufen des Forschungsprozesses eher geringe Aufmerksamkeit, obwohl auch sie für die Qualität der Ergebnisse entscheidend sind. Dieser Beitrag beschäftigt sich näher mit einem solchen Teilschritt: der fallbezogenen Edition von Datensätzen am Beispiel von Lebensverlaufsdaten.

Individuelle Lebensverlaufsdaten sind weitgehend standardisiert erhobene, aber in ihrer Struktur besonders komplexe empirische Daten. Daher ist hier der Aufwand für die Prüfung und eventuelle Korrektur der Daten auch nach der Erhebung besonders groß, und erst nach der Datenedition steht der eigentliche Analysedatensatz zur Verfügung. Im Falle der am Max-Planck-Institut für Bildungsforschung durchgeführten Lebensverlaufsstudien handelt es sich bei der Edition um explizit regelgeleitete, meist mehrjährige Prozesse, in denen die Daten fallweise überprüft werden. Hierbei stellen sich mitunter auch grundsätzliche theoretische Fragen. Anhand von Beispielen aus der jüngsten Teilstudie des

Lebensverlaufsprojektes soll im Folgenden auf zentrale Aspekte der Datenedition eingegangen werden.[1]

Das Projekt 'Lebensverläufe und gesellschaftlicher Wandel' wird unter der Leitung von Karl Ulrich Mayer seit 1979 durchgeführt, ab 1983 am Forschungsbereich Bildung, Arbeit und gesellschaftliche Entwicklung des Max-Planck-Instituts für Bildungsforschung. Die Datenbasis setzt sich im Wesentlichen aus einmaligen (nur in einer Teilstudie auch mehrmaligen), standardisierten Retrospektivbefragungen von Personen aus neun ausgewählten Geburtsjahrgängen zusammen, die systematisch über ihren bisherigen Lebensverlauf Auskunft geben (vgl. Brückner/Mayer 1998). In ihrer Gesamtheit erlauben diese Lebensverlaufsdaten für Deutschland detaillierte und historisch weit zurückreichende Analysen, die fast das gesamte 20. Jahrhundert erfassen.[2]

In der neuesten Teilstudie (vgl. Corsten/Hillmert 2001) werden die bis Ende der achtziger Jahre in der alten Bundesrepublik durchgeführten Befragungen durch Daten ergänzt, die die Lebensverläufe, insbesondere die Bildungs- und Erwerbsverläufe, der Geburtsjahrgänge 1964 und 1971 in Westdeutschland bis Ende der neunziger Jahre verfolgen. Im Gegensatz zu den bisherigen Lebensverlaufsstudien wurden auch in Deutschland lebende Ausländer in die Untersuchung einbezogen sowie bewusst zeitlich parallele Ausbildungs- und Erwerbsverläufe aufgenommen. Das Projekt wird in Zusammenarbeit mit dem Institut für Arbeitsmarkt- und Berufsforschung (IAB) in Nürnberg durchgeführt.[3] Die Haupterhebung der Daten erfolgte in den Jahren 1998 und 1999 in Kooperation mit dem Umfrageinstitut Infas (Infas 1999). Erfolgreich befragt werden konnte eine Stichprobe von rund 2.900 Personen. Die Interviews erfolgten zumeist in der CATI (computer-assisted telephone interviewing)-Technik. Seit 1999 werden die Daten im Max-Planck-Institut für Bildungsforschung geprüft und ediert.

2. Grundprobleme der Datenedition

Die Daten der einzelnen Lebensverlaufsstudien sind nach Lebensbereichen in verschiedenen Modulen abgelegt (zum Beispiel Schulgeschichte; Ausbildungsgeschichte; Erwerbsgeschichte; Partnerschaftsgeschichte). Dabei ist zwischen Personendatensätzen und Episodendatensätzen zu unterscheiden. Erstere enthalten im Wesentlichen Querschnittsanga-

---

1 Der Verfasser hat die Edition dieser Teilstudie seit Anfang 2000 geleitet. Für hilfreiche Kommentare zu diesem Beitrag danke ich Karl Ulrich Mayer, Heike Trappe sowie den anonymen Gutachtern.

2 Vgl. dazu http://www.mpib-berlin.mpg.de/de/forschung/bag/index.htm

3 In diesem Zusammenhang gefördert durch Mittel des Europäischen Sozialfonds.

ben (über die Zeit konstante Personenmerkmale), letztere nehmen für jede Person einzelne Episoden (zum Beispiel Ausbildungen oder Arbeitsverhältnisse) und deren Merkmale gesondert auf. Ziel der Lebensverlaufsstudien ist es, für jeden Lebensbereich möglichst präzise (monatsgenau) und vollständig die Sequenz einzelner lebensgeschichtlicher Abschnitte zu rekonstruieren.

Derart komplexe Daten sind fehleranfällig. Sie werden daher nach der Erhebung einem aufwändigen Editionsprozess unterzogen (vgl. Tölke 1982; Brückner et al. 1983). Hierbei werden alle Fälle im Hinblick auf fehlende oder unplausible Angaben geprüft. Um die Nachvollziehbarkeit des Editionsprozesses zu sichern, werden eine Reihe von Prinzipien zugrunde gelegt (siehe auch die Dokumentationen zur Edition in verschiedenen Teilstudien des Lebensverlaufsprojekts). Grundsätzlich gilt, dass nach Möglichkeit alle Unplausibilitäten aufzuklären und alle verfügbaren Informationen aus den Interviews auszuschöpfen sind.

Ein großer Teil des auftretenden Korrekturbedarfs ist nicht inhaltlich gelagert: hier handelt es sich um die Folgen von Steuerungsfehlern im Instrument, um Tipp- oder Hörfehler der Interviewer und ähnliches. Diese können oft einfach und mit hoher Sicherheit korrigiert werden, wie auch im nächsten Abschnitt deutlich werden wird. In einigen Fällen stellen sich aber stärker inhaltliche und auch theoretisch bedeutsame Probleme. Auf die wichtigsten soll hier kurz eingegangen werden.

(1) Ist eine inhaltlich orientierte Edition, die über die Korrektur erhebungstechnischer Fehler hinausgeht, überhaupt sinnvoll? Mögliche Fehler in den Daten fallen in der Prüfung auf, wenn die Angaben logisch inkonsistent sind oder nicht im Rahmen des Erwartbaren liegen. Prinzipiell besteht damit die Möglichkeit, dass durch Korrekturen an den Daten eine künstliche Konsistenz in den Angaben hergestellt, also ungewöhnliche Lebensverläufe ‚geglättet' werden könnten. Angaben der Befragten können unplausibel erscheinen, obwohl sie korrekt sind. Die Daten würden fehlerhaft, wenn hier eine Korrektur erfolgte.

In der Edition kann zum Beispiel das Verhältnis von Ausbildung und Erwerbstätigkeit unplausibel erscheinen, d.h. erworbene Qualifikationen und Tätigkeit einer Person passen offenbar nicht zusammen. Eine einfache Korrektur der Angaben in einer der beiden Variablen wäre ein heikler Eingriff, zumal es sich bei diesem Vergleich um einen Zusammenhang handelt, der selbst Gegenstand soziologischer Forschung ist (man denke an Themen wie ‚Bildungsrenditen' und ‚ausbildungs(in)adäquate Beschäftigung'). Für die Ergebnisse bestünde die Gefahr von Artefakten, wenn bereits im Prozess der Edition Ausbildungs- und Beschäftigungsniveau einander angepasst werden und dann in den späteren Analysen die hochgradige Übereinstimmung als Forschungsergebnis präsentiert wird.

Wie dieses Beispiel zeigt, ist ein Fehler aber immer in beide Richtungen - durch Korrektur und durch Unterlassen - möglich. Da die Informationen in den meisten Variablen der Studie als objektive Aussagen über die Realität (etwa formale Bildungsabschlüsse) und nicht als rein subjektive Wahrnehmungen verstanden werden, führen auch unterbleibende Korrekturen zu Fehlern in den Ergebnissen. Dies ist der Fall, wenn die in Analysen festgestellte Inkonsistenz eher auf lückenhaften Angaben als auf realen Abweichungen beruht, wenn also etwa vorhandene Ausbildungen nicht genannt oder nicht aufgenommen wurden. Ähnliches gilt, wenn sich die Befragten selbst (objektiv) falsch einer Kategorie zuordnen. In diesen Fällen würde es zu unrealistisch hoher Inkonsistenz kommen. Diese Art von Fehlern scheint besonders dann möglich, wenn die Befragten mit Konzepten konfrontiert werden, die nicht unbedingt Teil ihrer Lebenswelt sind - von Problemen sozialer Erwünschtheit beim Antwortverhalten einmal ganz abgesehen. Im konkreten Fall dürfte dies beispielsweise die Variable der beruflichen Stellung (mit Kategorien wie ‚Angestellte mit einfacher Tätigkeit' oder ‚Angestellte mit qualifizierter Tätigkeit') überdurchschnittlich betreffen. Auch wenn den Befragten für die einzelnen Kategorien jeweils Beispiele von Tätigkeiten vorliegen, dürfte ihnen diese eher abstrakte Skala weniger vertraut sein als etwa die Bezeichnung ihrer konkreten Ausbildungsabschlüsse.

Jede einzelne Editionsentscheidung stellt also eine Risikoabwägung dar: fehlerhafte Angaben sollen möglichst korrigiert, fehlerhafte Korrekturen möglichst vermieden werden.

(2) Wie lässt sich eine fehlerhafte Edition verhindern? Verschiedene Vorkehrungen gegen Fehler dieser Art sind möglich. Basis der Plausibilitätsprüfung (Wie lange dauerte eine bestimmte Ausbildung? Ist das Verhältnis von Ausbildung und Erwerbstätigkeit plausibel?) sollten nicht nur die empirischen Mittelwerte, sondern entsprechende Streuungsbereiche und auch mögliche Ausnahmen sein. Bei einer detaillierten Befragung wie der Lebensverlaufsstudie stehen zudem meistens Informationen aus anderen Variablen zur Verfügung, im vorliegenden Beispiel etwa nähere Angaben zur betreffenden Erwerbstätigkeit. Dies ist übrigens auch der Grund für die Entscheidung, die Edition *einzelfallweise* durchzuführen: ein solches Vorgehen erlaubt es, nicht nur alle Variablen innerhalb einer Episode zu berücksichtigen, sondern diese sowohl in einen zeitlichen Verlauf als auch in den Kontext verschiedener Lebensbereiche einzuordnen. Daneben sollten auch Sekundärinformationen wie institutionelle Vorgaben herangezogen werden (etwa gesetzliche Zulassungsregeln). Bei der Beurteilung der Glaubwürdigkeit einzelner Angaben können im Einzelfall Kriterien wie die Menge und Detailliertheit von Angaben, Sicherheit im Antwortverhalten oder die zu erwartende Kenntnis bestimmter Kategorisierungen berücksichtigt werden.

Grundsätzlich sollten Entscheidungen nicht ad hoc getroffen, sondern in der Editionsgruppe argumentativ abgewogen werden. Individuelle Entscheidungen der Bearbeiter sollen unterbleiben. Die Plausibilitätskontrolle dient zunächst auch nur dazu, bestimmte Fälle oder Fragen noch einmal genauer zu prüfen; aus ihr sollte keine automatische Korrekturvorschrift folgen, Eingriffe in die Daten sollten vorsichtig gehandhabt werden. Gibt es keinen hinreichenden Beleg für eine veränderte Kategorisierung, sollte die Angabe unverändert belassen werden. Bestehen Zweifel an der Richtigkeit der Angaben, lässt sich dies aber dadurch markieren, dass zusätzliche Angaben neben die Originalangaben treten (explizite Korrekturvariablen, als ungefähr kenntlich gemachte Zeitangaben und ähnliches). Auch eine eventuelle Unsicherheit der Informationen kann also ausgewiesen werden. Die Lebensverlaufsstudie enthält auch explizit subjektive Variablen (etwa die Frage nach den Gründen für den Wechsel eines Arbeitsverhältnisses), bei denen keine objektiven Plausibilitätskriterien angelegt werden. Auch hier können die Daten aber zum Beispiel im Hinblick auf Eingabe- und Übertragungsfehler geprüft werden.

(3) Schließlich könnte die Gefahr des ‚Kategorienfitting' bestehen: Werden die Angaben auch und gerade in der Editionsphase dem starren Erhebungsinstrument angepasst? Eine Kritik an der (quantitativen) empirischen Sozialforschung lautet, dass sie durch vorab definierte Kategoriensysteme entscheidende Tendenzen des sozialen Wandels übersieht. Über derart „blinde Empirie" schreiben etwa Beck/Beck-Gernsheim (1994: 32; 39): „Quantitative Methoden setzen Kategorisierungen, Gruppenbegriffsbildungen voraus (...). Eine sich individualisierende Gesellschaft entzieht sich aber diesen untersuchungstechnischen Standardisierungszwängen (...). Deshalb ist es für eine auf ihre technische Brillanz stolze Soziologie schwer, sich über ihren eigenen Schatten hinweg für die Fragen der sich individualisierenden Gesellschaft zu öffnen."

Bei der Edition sollte es aber nicht darum gehen, lückenhafte Angaben um jeden Preis zu füllen, unplausible Angaben anzupassen oder Antworten auf jeden Fall in die festgelegten Kategorien einzuordnen. Für die einzelfallorientiert arbeitende Edition spielen zwar Motive der *Praktikabilität* durchaus eine Rolle - etwa das grundsätzliche Ziel, die weitgehende Vollständigkeit der Daten für Analysezwecke herzustellen: eine sehr lückenhafte Datenbasis wäre nur bedingt verwendungsfähig. Dies gilt aber nicht absolut. Durch die Einzelfallprüfung kommt die Edition zudem bewusst dem Anspruch nach möglicher kategorialer Differenzierung entgegen, und die Hauptleitlinie bleibt die der *Qualität* der Daten im Sinne der Übereinstimmung mit der objektiven Realität bzw. der tatsächlichen Wahrnehmung der Zielpersonen. So kann die Tatsache, dass Informationen beim Interview nicht richtig zugeordnet werden konnten, gerade bei der Edition auffallen; solche Probleme zeigen sich insbesondere in der starken Besetzung von Restkategorien oder

auch in Rückfragen der Interviewten. Zum anderen können im Zuge der Edition Kategorienschemata verändert werden, um letztlich möglichst alle Antworten präzise zu erfassen.

3. Edition in der Lebensverlaufsstudie 1964/71

3.1 Editionsverlauf

Gerade unter Berücksichtigung der eben genannten Anforderungen ist die Edition von detaillierten Lebensverlaufsdaten bei großen Stichproben eine umfangreiche Aufgabe, die im Projekt nur arbeitsteilig bewältigt werden kann. Mit den unmittelbaren Prüfungen und Korrekturen ist in der jüngsten Lebensverlaufsstudie insbesondere ein Team von bis zu zwölf studentischen Mitarbeiterinnen und Mitarbeitern betraut gewesen. Für die Edition wurde ein differenziertes Regelwerk konstruiert, das sukzessive angepasst wurde. Auftretende Inkonsistenzen und neue Probleme wurden in der Arbeitsgruppe diskutiert und entschieden. Die einzelnen Bearbeiter erhielten Supervision. Schließlich wurden alle Regeln und auch alle (Einzelfall-)Entscheidungen ausführlich dokumentiert (Spengemann/Hillmert 2002).

Die Datenedition hat sich hier in mehreren Phasen über einen Zeitraum von etwa dreieinhalb Jahren erstreckt. Die folgenden Absätze geben eine kurze, vereinfachte Beschreibung dieser Editionsphasen.

Eine *Voredition* ausgewählter Interviews begann bereits kurz nach Beginn der Erhebung. Der Schwerpunkt lag dabei in erster Linie auf der Beurteilung der Interviewer und erst in zweiter Linie auf der Edition der Daten. Fehler und problematische Stellen des Interviews sollten so genau wie möglich protokolliert werden. Grundlage waren die Biographieschemata (Auflistung der Lebensverlaufsepisoden in standardisierter Form) und Interviewerprotokolle (zusätzliche handschriftliche Notizen der Interviewer). Außerdem standen in den meisten Fällen Tonbandaufzeichnungen der Interviews (mit Einverständnis der Befragten) zur Verfügung.

Die *Erstedition* der 2911 Fälle begann im Januar 1999 nach einer Schulung der Editionsgruppe. Die Arbeitsmaterialien wurden erweitert: es zählten hierzu eine erste Version des Editionshandbuchs, die Variablenliste, das Computerprogramm für die Erhebung (um die Steuerung nachvollziehen zu können) sowie der Rohdatensatz in der Form von Einzelfallprotokollen (also alle Angaben eines Falles als Papierausdruck) und, falls vorhanden, die Ergebnisse der Voredition. Die Editeure waren gehalten, keine eigenmächtigen Entscheidungen zu treffen, sondern Probleme zuerst mit der Supervision zu besprechen. Die passenden Editionsregeln wurden dann noch einmal erläutert, oder der Fall wurde in einer Editionssitzung vorgestellt beziehungsweise der Projektleitung zur Einzelfallentscheidung

vorgelegt. Aus Gründen der inhaltlichen Fokussierung und der verfügbaren Arbeitskapazität konzentrierte sich die Edition auf die Hauptmodule (im Wesentlichen Schul- und Berufsausbildungen, Arbeitslosigkeit, Erwerbstätigkeiten und die dazwischen liegenden Lücken, sowie persönliche Angaben), dabei wurden aber auch die Informationen aus den anderen Modulen genutzt. Eine Reihe von Fällen wurde zur Überprüfung, Korrektur und Anpassung an die inzwischen deutlich veränderten Editionsregeln wiederholt ediert.

Eine wichtige Informationsmöglichkeit war anschließend die telefonische *Nachrecherche* bei den Zielpersonen. Falls nötig, wurden sie noch einmal telefonisch kontaktiert und um klärende oder ergänzende Informationen gebeten. Die telefonische Nachrecherche startete nach einer Schulung der damit betrauten Mitarbeiter im April 2000. Zu diesem Zeitpunkt waren bereits durch die Erstedition rund 700 Fälle dafür vorgesehen. Im Rahmen der Endedition kamen weitere Fälle hinzu, so dass schließlich 1.272 Fälle als Nachrecherchefälle deklariert waren. Davon wurden 110 Fälle nicht bis zum festgelegten Ende der Nachrecherche (Dezember 2001) erreicht, bei einigen wenigen Fällen schienen weitere Kontaktversuche nicht sinnvoll. Ein Teil der Fälle musste lediglich wegen der fehlenden Note eines Ausbildungsabschlusses nachbefragt werden (Steuerungsfehler im Befragungsinstrument der Haupterhebung). Alle Personen sollten zunächst telefonisch kontaktiert werden, und das Nachrecherche-Interview sollte möglichst telefonisch geführt werden. Vom Erhebungsinstitut wurde die für diesen spezifischen Zweck notwendige Kontaktinformation zur Verfügung gestellt. War keine Telefonnummer vorhanden oder diese nicht mehr korrekt, mussten vorab Nachforschungen durchgeführt werden; war dies über keines der verwendeten Telefonnummernverzeichnisse möglich, wurden die Personen zwecks Terminanfrage angeschrieben. In 141 Fällen wurden weitere Auskünfte explizit verweigert oder Terminanfragen nicht beantwortet, so dass insgesamt 1.011 Nachrecherchen erfolgreich durchgeführt werden konnten. Bei der Nachrecherche ging es in zwei Drittel der Fälle um fehlende Angaben in ein oder zwei Modulen bzw. Unklarheiten in maximal drei Episoden, es handelte sich also im Wesentlichen um gezielte Ergänzungen.

Die *End- bzw. Hauptedition* wurde zusammen mit der telefonischen Nachrecherche begonnen und für alle Fälle durchgeführt. Zusätzlich zu den bislang vorhandenen Materialien stand jetzt auch der Fragebogen in einer ausformulierten Papierversion zur Verfügung, die aus dem Programmcode des Erhebungsinstruments erstellt worden war. Das Editionshandbuch wurde speziell für die Endedition ein weiteres Mal überarbeitet, insbesondere im Abschnitt zum Modul Berufsausbildung. Aber auch nach Beginn der Endedition wurden Editionsregeln weiter angepasst und vervollständigt.

Ende 2002 wurde die Edition durch die automatisierte, modulweise Datenprüfung abgeschlossen. Daneben gab es zu jedem Zeitpunkt der Edition Einzelfallentscheidungen. Sie

wurden getrennt für die verschiedenen Phasen dokumentiert, da im Zuge der Endedition die eine oder andere Entscheidung aus einer früheren Phase rückgängig gemacht werden musste. Wurde beispielsweise eine Editionsregel mit weitreichenden Folgen neu erstellt, so mussten alle Fälle, die bis zu diesem Zeitpunkt bereits bearbeitet waren, entsprechend korrigiert werden. Dies zeigt, dass sich die Edition im fortgeschrittenen Stadium nicht nur mit den Rohdaten, sondern immer wieder mit eigenen früheren Entscheidungen auseinandersetzen musste. Nach Abschluss der Edition weist die Dokumentation nur noch die Einzelfallentscheidungen auf, die erhalten blieben oder während der Endedition neu getroffen wurden.

Für die Endedition mit telefonischer Nachrecherche, die die bei weitem umfangreichste Phase der fallorientierten Edition war, wurde eine Statistik zum Zeitaufwand bei der konkreten Fallarbeit geführt. Bei einem mittleren Zeitaufwand von etwa 100 Minuten und etwa 2.900 Fällen ergibt sich daraus ein Gesamtvolumen von etwa 5.000 Arbeitsstunden. Der Arbeitsaufwand für die Edition insgesamt - inklusive Schulungen, Regelkonzeption, Interviewvorbereitung, Einzelfallentscheidungen, Vercodungen, Organisation usw. - hat ein Mehrfaches dieser Zeit beansprucht.

3.2 Vorgehen bei der Datenedition: einige Fallbeschreibungen

Im Zuge der Edition wurden also alle Fälle einzeln auf Basis verschiedener Informationsquellen wie dem Rohdatensatz aus der Befragung, Interviewerprotokollen und zumeist der Tonbandaufzeichnung des Interviews geprüft. Gerade mit Hilfe des Tonbandes konnte eine Reihe von Unplausibilitäten relativ einfach geklärt werden, wenn daraus beispielsweise eindeutig hervorging, dass der Interviewer zwei Zahlen vertauscht hatte. Daneben wurden auf dem Tonband auch zusätzliche Informationen wie Interaktionen zwischen Interviewer und Befragtem (Erläuterungen, Rückfragen etc.) festgehalten. Hinzu kam als wichtige Informationsmöglichkeit die telefonische Nachrecherche. Da diese in einem zeitlichen Abstand von bis zu drei Jahren gegenüber der Hauptbefragung erfolgte, kann die Nachrecherche ihrerseits mit einigen Konsistenz- und Erinnerungsproblemen (Reimer 2001) behaftet gewesen sein, auf die hier nicht weiter eingegangen werden kann. Schließlich wurden Informationen aus externen Quellen (etwa aus Handbüchern zur Berufskunde oder dem Internet) genutzt.

Die Konsistenzprüfung war nicht die einzige Aufgabe der Edition, aber doch ein wesentlicher Teil. Mögliche Inkonsistenzen wurden bereits bei der Erhebung reduziert, da während des Interviews eine automatische (programmgesteuerte) Konsistenzprüfung zwischen bestimmten Angaben stattfand. Diese Prüfung betraf insbesondere zeitliche Unplausibilitäten (Beispiel: das Ende einer Episode lag zeitlich angeblich vor ihrem Beginn) und Lücken. In solchen Fällen wurde der Interviewer entweder unmittelbar zu einer Nach-

frage aufgefordert, oder es wurde am Ende des Interviews noch einmal gezielt nach Aktivitäten in den Lückenzeiträumen gefragt.

Ursachen für verbleibende Inkonsistenzen in den Daten können Instrumentfehler wie eine fehlerhafte Steuerung oder eine unterbliebene Präzisierung bzw. Differenzierung von Kategorien, Fehler der Interviewer, eine im Sinne der vorgegebenen Kategorien falsche Selbsteinschätzung der Befragten, Erinnerungsfehler oder bewusst falsche Angaben sowie reale Abweichungen gewesen sein.

Die folgende Übersicht (Tabelle 1) zeigt wesentliche Arten von Inkonsistenzen im Hinblick auf die modulare Struktur der Daten:

Tabelle 1: Dimensionen der Konsistenzprüfung und mögliche Inkonsistenzen in den Daten (Beispiele)

	Innerhalb von Modulen	**Zwischen Modulen**
Synchron	(a) Variablen innerhalb einer Episode; z.B. Ausbildungsstätte (Fachschule) entspricht nicht dem Abschluss (Fachhochschuldiplom) (b) Parallele Episoden; z.B. mehrere parallele Ausbildungen in Vollzeit	Parallelität von Zuständen; z.B. längere Ausbildung im Ausland, obwohl im Inland gewohnt
Diachron	Mehrfachausbildungen und Aufbauqualifikationen; z.B. Meisterqualifikation ohne vorherige Berufsausbildung	Reihenfolge (und Niveau) von Aktivitäten; z.B. hochqualifizierte Tätigkeit ohne Ausbildung

Die Edition hat sich dabei auf die Module konzentriert, die auch im Zentrum der Forschungen am Institut stehen, nämlich Ausbildungen und Erwerbstätigkeiten. Das Vorgehen soll zunächst durch die Darstellung zweier konkreter Fälle erläutert werden. Diese Lebensverläufe können nicht inhaltlich ‚repräsentativ' für die gesamte Stichprobe sein, und sie sind nicht in ihrer ganzen Komplexität dargestellt; es kommen hier aber für die Datenedition typische Anforderungen zum Ausdruck:

Fall A. (Jahrgang 1971)

Daten vor Edition

Frau A., die 1977 eingeschult wurde, beendet im Januar 1985 die Sonderschule nach der achten Klasse mit dem einfachen Hauptschulabschluss. Sie absolviert von September 1986 bis September 1987 ein Berufsvorbereitungsjahr und beginnt 1986 eine Bäckerlehre, an deren genauen Anfangszeitpunkt sie sich nicht erinnert. Diese Ausbildung sagt ihr aber nicht zu, zudem hat sie gesundheitliche Probleme. Sie bricht die Bäckerlehre

ab, da sie eine bessere Ausbildungsstelle als Hauswirtschaftshelferin findet. Der fehlende Zeitraum zwischen dem Ende der allgemeinbildenden Schule im Januar 1985 und dem Beginn des Berufsvorbereitungsjahres im September 1986 kann nur durch die Informationen aus einer Lückenabfrage am Ende des Interviews ergänzt werden: Frau A. ist von Januar 1985 bis Juni 1986 krank und besucht von Mai 1986 bis September 1987 nochmals eine allgemeinbildende Schule.

Aufgabenstellung für die Edition
Der Abschluss der Sonderschule mit dem einfachen Hauptschulabschluss nach acht Schuljahren ist zu überprüfen. Die Überschneidung zwischen der Krankheitsphase von Januar 1985 bis Juni 1986 und dem ebenfalls nur aus der Lückenabfrage bekannten weiteren Besuch einer allgemeinbildenden Schule von Mai 1986 bis September 1987 ist ebenso unplausibel wie die Parallelität dieses Schulbesuchs zum Berufsvorbereitungsjahr.

Bearbeitung durch Edition und Nachrecherche
Die Edition zieht zunächst den Interviewmitschnitt zur Klärung dieser Inkonsistenzen heran und kann feststellen, dass Frau A. im Zeitraum von Mai 1986 bis September 1987 keine weitere allgemeinbildende Schule, sondern von September 1986 bis September 1987 tatsächlich ein Berufsvorbereitungsjahr absolviert hat. Somit kann der als sogenannte Lückenaktivität erfasste weitere Schulbesuch gestrichen werden. Zur Klärung der anderen offenen Fragen muss eine Nachbefragung durchgeführt werden: Frau A. beendet die Sonderschule im April 1986 nach der neunten Klasse nicht mit einem Hauptschulabschluss, sondern mit dem Sonderschulabschluss. Den einfachen Hauptschulabschluss bekommt sie erst im Juli 1989 mit dem Abschluss der Ausbildung zur Hauswirtschaftshelferin zuerkannt. Vor dem Beginn des Berufsvorbereitungsjahres im September 1986 macht Frau A. ab Mai 1986 ein zweimonatiges Betriebspraktikum in einer Bäckerei ‘um diesen Beruf kennen zu lernen‘. Direkt im Anschluss beginnt sie im Juli 1986 die Bäckerlehre. Sie bricht diese Lehre nach zwei Monaten ab und beginnt anschließend das Berufsvorbereitungsjahr. Die in der Lückenabfrage erhobene Krankheit war lt. Frau A. keine dauerhafte Phase.

Fall B. (Jahrgang 1964)

Daten vor Edition
Herr B. schließt im April 1983 eine Maurerlehre ab und wird im Anschluss von seinem Ausbildungsbetrieb als Maurer übernommen. Er arbeitet bis Mai 1992 (ohne Unterbrechung) als Facharbeiter in diesem Betrieb. Andererseits ist Herr B. nach Angaben aus dem Interviewerprotokoll aber in den Jahren 1983/84, 1984/85 und 1985/86 jeweils in den Monaten Dezember bis April arbeitslos. Für den Zeitraum zwischen der im April 1983 beendeten Maurerausbildung und April 1986 gibt er in der Lückenabfrage drei Erwerbstätigkeitszeiträume an: von April bis Dezember 1983, von April bis Dezember 1984 und von April bis Dezember 1985. Herr B. hat von April 1986 bis Juni 1986 neben seiner Erwerbstätigkeit als Maurer eine weitere Ausbildung zum Sprengmeister absolviert. Sein Vater war in diesem Beruf selbstständig, und um ‚dessen Betrieb übernehmen zu

können‘, benötigte Herr B. eine ‚Lizenz‘. Diese Ausbildung fand in Teilzeitform an einer Berufsfachschule statt, und Herr B. beendete sie mit dem Meistertitel.

Aufgabenstellung für die Edition

Die Phasen der Arbeitslosigkeit in den Monaten Dezember bis April der Jahre 1983/84, 1984/85 und 1985/86 liegen innerhalb des über neun Jahre andauernden Erwerbstätigkeitszeitraumes als Maurer von 1983 bis 1992. Gleiches gilt für die nur aus der Lückenabfrage bekannten Erwerbstätigkeitsphasen April bis Dezember 1983, April bis Dezember 1984 und April bis Dezember 1985. Besteht der neunjährige Erwerbszeitraum also aus mehreren von Arbeitslosigkeit unterbrochenen Abschnitten oder handelt es sich um unterschiedliche Erwerbsverhältnisse? Die Zuordnung der Qualifikation zum Sprengmeister als Ausbildung ist zu prüfen. Die Berufsfachschule ist hier als Ausbildungsstätte ebenso in Frage zu stellen wie die Meisterqualifikation.

Bearbeitung durch Edition und Nachrecherche

Zur Klärung kann die Edition Interviewprotokoll und Tonband nutzen: Die im Interviewprotokoll festgehaltenen Phasen saisonbedingter Arbeitslosigkeit in den Wintermonaten von 1983 bis 1992 werden durch das Abhören des Tonbandes teilweise bestätigt: Herr B. ist bis einschließlich 1985/86 jeweils von Dezember bis März arbeitslos. Er muss die neunjährige Erwerbstätigkeit als Maurer in diesen Zeiträumen unterbrechen. Die Erwerbstätigkeit ist also zwischen Mai 1983 und Mai 1992 in vier Phasen aufzuteilen: April bis Dezember 1983, April bis Dezember 1984 und April bis Dezember 1985 und April 1986 bis Mai 1992. Das Abhören des Tonbandes bestätigt auch, dass die drei in der Lückenabfrage für die Monate April bis Dezember 1983-85 aufgenommenen Erwerbstätigkeitszeiträume mit den ersten drei dieser Erwerbstätigkeitsphasen (April bis Dezember in den Jahren 1983-85) identisch und deshalb zu streichen sind. Genauere Informationen zu den vier Erwerbstätigkeitsphasen müssen durch die Nachbefragung ergänzt beziehungsweise komplett nacherhoben werden: Herr B. ist von Mai 1983 bis November 1983, von April 1984 bis Dezember 1984, von April 1985 bis Dezember 1985 und von April 1986 bis Mai 1992 jeweils als Maurer in der gleichen Firma beschäftigt. Für diese Phasen können jeweils die fehlenden Informationen ergänzt werden. Mit Hilfe der berufskundlichen Informationen des Arbeitsamtes kann ferner festgestellt werden, dass es sich beim Sprengmeister um einen Anlernberuf handelt, für den ein staatlich anerkannter Lehrgang oder eine erfolgreiche Sachkundeprüfung vorzuweisen ist. Eine Meisterqualifikation im Sinne der Edition wird nicht erreicht, sondern lediglich die Berechtigung zur Ausübung bestimmter Tätigkeiten. Die Ausbildung wird als berufliche Weiterbildung erfasst.

An die Betrachtung einzelner Fälle schließt sich unmittelbar die Frage an, inwieweit sich Merkmalsverteilungen und Zusammenhänge in der gesamten Stichprobe durch ein solches Editionsverfahren verändern. Das Ausmaß dieser Veränderungen variiert natürlich je nach Variable bzw. inhaltlicher Problemstellung. Im Folgenden sollen solche Veränderun-

gen anhand von Beispielen aus den beiden für die Studie wohl wichtigsten Modulen, den Ausbildungs- und Erwerbsgeschichten, dargestellt werden.

3.3 Veränderungen von uni- und bivariaten Merkmalsverteilungen im Zuge der Edition

In die Unterschiede zwischen Merkmalsverteilungen, die auf einem unterschiedlichen Datenstand beruhen, gehen mehrere Veränderungsdimensionen ein. Im Zuge der Edition wurden nicht nur Angaben in allen einzelnen relevanten Variablen, sondern auch die Struktur der Ausbildungs- und Erwerbsgeschichten (Episodenanzahl) verändert. Bei diesen Episoden handelt es sich nicht in jedem Fall um ganze Ausbildungen oder Erwerbsverhältnisse; vielmehr sind es häufig Teilabschnitte. Auch bei gleicher Episodenstruktur können sich aber die Werte in den einzelnen Variablen verändern.

Die Tabellen im Anhang geben anhand einiger Beispiele Auskunft über Verteilungsänderungen im Zuge der Edition. Verglichen wird jeweils der Datenstand nach dem Abschluss der Einzelfalledition mit jenem vor Beginn der Edition. Tabelle 2 bezieht sich auf die Variable ‚Ausbildungsabschluss' im Modul Ausbildungsgeschichte, Tabelle 3 auf die Variable ‚differenzierte berufliche Stellung' im Modul Erwerbsgeschichte, wobei jeweils nur ausgewählte (stärker besetzte) Kategorien ausgewiesen werden. Die Werte für die Spalte ‚nach Edition' beziehen sich bei der Variable ‚berufliche Stellung' auf die im Zuge der Edition eingeführte Korrekturvariable - neben diesem Vorschlag der Edition blieben hier also die Originalangaben im Datensatz erhalten. In beiden Tabellen wird ferner unterschieden zwischen einer Betrachtung aller Episoden einerseits und nur jeweils der ersten genannten Episode andererseits. Letztere lässt sich auch als Verteilung von Personen interpretieren. Schließlich werden Veränderungen in der Episodenstruktur auf der Individualebene angegeben. Es zeigt sich, dass sich bei mehr als einem Fünftel der Befragten bereits die Anzahl der angegebenen Ausbildungs- oder Erwerbsepisoden im Zuge der Edition verändert hat.

Im Vergleich der Datensätze mit unterschiedlichem Editionsstand lassen sich für die Verteilungen der beiden Variablen folgende Unterschiede erkennen:

(1) Variablen unterscheiden sich im Ausmaß der Veränderung. Für die Variable Ausbildungsabschluss zeigen sich insgesamt eher größere, für die beruflichen Positionskategorien eher geringere Verschiebungen in der relativen Besetzung. Dies dürfte auch auf die bessere Abgrenzbarkeit der einzelnen Ausbildungskategorien zurückgehen.

(2) Nicht alle Kategorien sind gleich betroffen. Gerade bei den Ausbildungsabschlüssen geht ein Teil der Veränderungen auf die Präzisierung einzelner Kategorien zurück. So waren in der Haupterhebung als Ausbildungskategorien sowohl ‚Meister/Techniker' als

auch ‚Fachschulabschluss' vorhanden. Dies führte zu Überschneidungen. Mit der Präzisierung der Kategorie ‚Fachschulabschluss' (‚... ohne Meister/Techniker') wurde dieses Problem im Zuge der Edition bereinigt. Auch bei den beruflichen Positionen gibt es bei einzelnen Kategorien zum Teil große relative Veränderungen: der Anteil der Episoden, die höheren Leitungsfunktionen zuzurechnen sind, nimmt beispielsweise deutlich ab.

(3) Grobe Kategorien (Zusammenfassungen) und einfache Maßzahlen sind offenbar prozentual weniger betroffen als einzelne Kategorien (ähnliches gilt für Mittelwerte bei anderen Variablen). Hier fallen Abweichungen in einzelnen Kategorien entweder nicht ins Gewicht, oder sie heben sich tendenziell auf. Fasst man beispielsweise alle Lehrabschlüsse zusammen, so erhält man hierfür vor Edition 68,2 und nach Edition 68,0 Prozent, für alle (Fach-)Hochschulabschlüsse 9,6 und 8,8 Prozent (jeweils erste Episode). Ähnliches gilt, wenn man bei der beruflichen Stellung alle Arbeiter (38,4 und 37,3 Prozent) oder alle Angestellten zusammenfasst (55,8 und 53,3 Prozent). Bei einzelnen Kategorien sowie bestimmten Zusammenhängen kann die Abweichung natürlich deutlich größer sein. Extrem große Unterschiede können sich auf der Ebene des Einzelfalles ergeben, wenn etwa einzelne Episoden vollständig gelöscht werden.

In einem weiteren Schritt werden nun Zusammenhänge zwischen diesen beiden Variablen analysiert: betrachtet wird das Qualifikationsniveau der Befragten in Abhängigkeit von ihrer beruflichen Stellung. Selbst dieses einfache Beispiel setzt voraus, die Längsschnittstruktur beider Module zu berücksichtigen. Sowohl das Qualifikationsniveau als auch die berufliche Stellung sind zeitveränderlich, so dass auch die Kombination beider Variablen im Zeitverlauf zu analysieren ist. Basis sind hier die Episoden der Erwerbsgeschichte, für die jeweils das in diesem Zeitraum aktuell vorhandene Qualifikationsniveau bestimmt wird; sollte sich die Qualifikation während einer Erwerbsepisode verändert haben, wird diese entsprechend geteilt. Da die beiden Module in ihrer genauen Zeitstruktur aufeinander bezogen werden, können sich nun zusätzlich Änderungen auswirken, die die Edition an den (monatsgenauen) Zeitangaben sowohl der Ausbildungs- als auch der Erwerbsepisoden vorgenommen hat; auf diese komplexen Zusammenhänge kann hier allerdings nicht näher eingegangen werden.

Tabelle 4 zeigt als Resultat aller Einflüsse die Veränderungen in der Verteilung der Ausbildungsabschlüsse nach beruflicher Stellung, wieder für alle Erwerbsepisoden und die jeweils erste genannte Erwerbsepisode. Erneut wird deutlich, dass die Veränderung nicht alle Kategorien gleichermaßen betrifft. Es gibt teilweise beträchtliche Verschiebungen in den Qualifikationsverteilungen der einzelnen beruflichen Positionen; sie können in der Größenordnung von 10 bis 20 Prozentpunkten liegen, etwa beim Anteil abgeschlossener Ausbildungen bei den qualifizierten Angestellten.

Was bedeutet dies für inhaltliche Analysen, die über die Betrachtung einzelner Kategorien hinausgehen? Definiert man adäquate Ausbildungsniveaus für die einzelnen Positionen wie folgt: ‚keine abgeschlossene Ausbildung' für un- und angelernte Arbeiter sowie einfache Angestellte; ‚abgeschlossene Ausbildung' für Fach- und Vorarbeiter sowie qualifizierte Angestellte; ‚(Fach-)Hochschulstudium' für hochqualifizierte Angestellte; ‚Berufsausbildung oder Studium' für Angestellte mit höheren Leitungsfunktionen, kann man den Anteil der (in)adäquat Beschäftigten in diesen Gruppen insgesamt grob bestimmen. Es zeigt sich, dass die Edition zwar in der Gesamtverteilung zu keiner qualitativen Veränderung des klaren (und i.d.R. monotonen) Zusammenhangs zwischen Position und Qualifikationsniveau in diesen Gruppen führt. Offensichtlich verändert sich aber die Stärke dieses Zusammenhangs, und er wird noch deutlicher. So erhält man vor Edition einen Anteil von 71,7 Prozent ausbildungs(niveau)adäquater Erwerbsverhältnisse. Nach Edition ergibt sich einen Anteil von 81,2 Prozent, wobei jeweils die erste Erwerbsepisode zugrunde gelegt wird. Die Anteilsdifferenz beträgt also knapp 10 Prozentpunkte.

Kommt man also mit unedierten Daten zu den gleichen inhaltlichen Schlussfolgerungen wie mit edierten Daten? Das hängt sicherlich von den zugrunde gelegten Hypothesen ab. Für beide Datenstände lässt sich eine einfache Aussage der Art ‚Die Mehrheit der Beschäftigten in den (ausgewählten) Arbeiter- und Angestelltenpositionen verfügt über das entsprechende Qualifikationsniveau' treffen. Deutlicher wird die Relevanz der Edition, wenn die Niveauunterschiede genauer betrachtet und möglichen Entwicklungen in der Grundgesamtheit gegenübergestellt werden: eine Aussage wie ‚der Anteil der adäquat Qualifizierten hat in einem (kurzen) Zeitraum um rund 10 Prozentpunkte zu- oder abgenommen' wäre sicherlich ein berichtenswertes Ergebnis, wenn er sich auf reale soziale Veränderungen zurückführen ließe. Die Entscheidung, ob und inwieweit eine detaillierte Datenedition durchführt wird, kann also solche Ergebnisse direkt beeinflussen.

3.4 Qualitätsverbesserung durch die Edition

Neben Konsistenzprüfungen im engeren Sinn konnte die Edition eine ganze Reihe weiterer Beiträge zur Verbesserung der Datenqualität in der vorliegenden Studie leisten. Hierzu zählen die Korrektur von Eingabefehlern und das Schließen von Lücken in den Angaben. Solche Lücken gehen nicht nur auf das individuelle Interviewer- bzw. Antwortverhalten, sondern auch auf mögliche Steuerungsfehler im Instrument der Ausgangserhebung zurück. Diese konnten so nachträglich korrigiert werden; ansonsten galt die genaue Fragebogenstruktur der Erhebung aber auch für Edition und Nachbefragung.

Schließlich hat die Edition die notwendige Ausdifferenzierung von Kategorien ermöglicht. In der Ausbildungs-Vercodung etwa wurden in einem separaten Arbeitsschritt unter anderem alle Fälle bearbeitet, die in der Variable Ausbildungsabschluss den Code der

Restkategorie und somit in einer Folgefrage offene Angaben enthielten. Ein Teil dieser Angaben konnte nachträglich einer der vorhandenen Kategorien zugeordnet werden. In vielen Fällen mussten aber aufwändige inhaltliche Recherchen angestellt werden, um die Nennungen präzise erfassen zu können. Bei der Variable Ausbildungsabschluss gab es vor Beginn der Edition 708 mal die Angabe ‚Anderer Ausbildungsabschluss' (= 21,0 Prozent der gültigen Angaben; vergleiche Tabelle 2), nach Edition nur noch 9 mal (0,2 Prozent). Dabei wurden neue Kategorien für Ausbildungsabschlüsse geschaffen, die bei der Erhebung nur unzureichend berücksichtigt worden waren (unter anderem verschiedene schulische Abschlüsse, Abschlüsse im Ausland, Berufsakademien). Das Kategoriensystem wurde also auf Basis der empirischen Angaben weiterentwickelt.

Die Kategorisierungen nach der Edition sind somit zum Teil erheblich differenzierter als vorher geworden. Genau dies aber kann ‚Kategorienfehler', wie sie in der oben zitierten Kritik angesprochen wurden, entgegenwirken. Trifft ein zu grobes oder veraltetes Kategoriensystem auf eine veränderte soziale Realität, so wird sich dies in der Form von Zuordnungsproblemen ausdrücken. Die Edition hat die Möglichkeit, eben diese Probleme sichtbar zu machen, darauf zu reagieren und adäquatere Kategorien zu finden. Dabei bleibt diese Eingriffsmöglichkeit nicht wie oft üblich auf Projektphasen beschränkt, die der Haupterhebung vorgelagert sind (Vorstudien, Pretests). Eine stillschweigende, verzerrende Anpassung der Antworten an ein unflexibles Kategoriensystem wird so verhindert.

Zusammenfassend lassen sich zahlreiche Belege dafür angeben, dass die Edition durch systematische Prüfung und gezielte Nachbefragung die Datenbasis ergänzen und deutlich verbessern konnte. Es ist daher durchaus legitim, Unterschiede zwischen edierten und nicht-edierten Daten im Wesentlichen als korrigierte Fehler zu interpretieren und diese dann auf Eigenschaften von Personen oder Interviewern zurückzuführen (vgl. Brückner 1995). Eine bisher noch nicht eingesetzte Möglichkeit, die Verbesserung der Datenqualität auch zu quantifizieren, wäre die systematische Gegenüberstellung von Daten unterschiedlichen Editionsstandes mit als valide angesehenen externen Informationen.[4]

4 Die Daten der jüngsten Lebensverlaufsstudie erlauben zum Teil die Verknüpfung mit prozessproduzierten amtlichen Längsschnittdaten (Individualinformationen aus der Beschäftigtenstatistik; mit Zustimmung der Befragten). Der Aufbau eines solchen Analysedatensatzes ist gegenwärtig in Arbeit, unter anderem als Teil eines Dissertationsprojektes. Während Vergleiche einfacher Querschnittsverteilungen auch mit vorhandenen Daten relativ unproblematisch möglich sind (Hillmert/Kröhnert 2001), sind individuelle Längsschnittvergleiche auf diese zusätzliche Datenquelle angewiesen.

Gerade in der Edition zeigen sich an verschiedenen Stellen auch Verbesserungsmöglichkeiten für zukünftige Datenerhebungen. So sollten für inhaltlich wichtige, aber eher abstrakte Konzeptionen und Skalen wie die berufliche Stellungen gezielt (weitere) Zusatzinformationen (etwa spezifische Tätigkeitsmerkmale) erhoben werden, die eine spätere Plausibilitätsprüfung erleichtern.

4. Fazit: Effekte und Effizienz der Edition

Für die eingangs geschilderte Systematik des Forschungsprozesses bleibt festzuhalten, dass durch systematische Edition qualitative Verbesserungen der Datenbasis auch und gerade in einer Phase möglich sind, in der die eigentliche Datenerhebung bereits abgeschlossen ist. Dies trägt dazu bei, den Gesamtfehler zu minimieren, solange die Konzentration auf die Edition nicht zu Lasten anderer Teilschritte geht. Gleichzeitig wird dabei von der linearen Struktur des Forschungsprozesses, die durch das idealtypische Modell nahegelegt wird, zugunsten einer Wiederholung beziehungsweise Korrektur einzelner Projektschritte und -entscheidungen abgewichen.

Lohnt sich der Aufwand einer solchen Datenedition? Mit Sicherheit verbessert sich die Datenqualität nicht nur im Sinne der Praktikabilität, sondern auch der ‚Wahrheitsähnlichkeit'. Ob der Ressourceneinsatz gerechtfertigt ist, bleibt natürlich eine Wertentscheidung, die sich darauf gründet, was den jeweiligen Forschern ein bestimmtes Ausmaß an Fehlerreduktion wert ist und wie detailliert soziale Prozesse abgebildet werden sollen. In jedem Fall muss mit Validitätsproblemen gerechnet werden, wenn auf den Schritt der Edition verzichtet wird. In der Lebensverlaufsstudie wird sicherlich überdurchschnittlich viel Wert auf die Datenedition gelegt, allerdings sind die meisten standardisierten Befragungen auch weniger komplex.

Wie groß sind die substantiellen Veränderungen infolge der Edition? Beim Vergleich grober Kategorien von Merkmalen zeigen sich eher geringe Veränderungen über die Editionsschritte hinweg als bei einzelnen, detaillierteren Kategorien. Man könnte nun argumentieren, dass eine an einfachen Anteilen bzw. Mittelwerten orientierte Analyse der Daten offensichtlich der geringsten Verfälschung unterliege und deshalb vorzuziehen sei. Dies ist aber nur ein Scheinargument, denn dann bräuchte man den Aufwand der Datenedition nicht betreiben. Außerdem unterliegen gerade auch einfache Verteilungsparameter der Stichprobenselektivität bei der Datenerhebung. Wenn man davon ausgehen kann, dass die Edition die Datenqualität tatsächlich deutlich verbessert, dann *sollten* derart aufwendig erhobene und bearbeitete Daten auch differenziert ausgewertet werden, sei es mit komplexeren quantitativen Verfahren, sei es stärker qualitativ bis hin zu individuellen Fallanalysen.

Auf jeden Fall sollte aber der Editionsprozess stärker als bisher in die inhaltlichen Analysen einbezogen werden, was die Vorstellung eines ‚in jeder Hinsicht korrekten Datensatzes' am Ende einer ergebnisneutralen Edition etwas relativiert. Prinzipiell ist für jede inhaltliche Fragestellung zu klären, ob und inwieweit sich die Edition der Daten auf den Erklärungsgegenstand (Verteilungen, Zusammenhänge) ausgewirkt hat oder haben könnte. Hierfür steht die umfangreiche Dokumentation der Edition zur Verfügung, die insbesondere für spätere externe Datennutzer wichtig sein dürfte. Die Berücksichtigung der Edition ist besonders dann wichtig, wenn der Forschungsgegenstand zumindest teilweise mit Verfahren und Rationalitäten übereinstimmt, die bereits die Phase der Edition bestimmt haben, das heißt wenn interessierende Zusammenhänge auch als Plausibilitätskriterium für die Prüfung und somit als Basis für Veränderungen gedient haben.

Für solche Zusammenhänge wiederum dürfte die Bestimmung absoluter Niveaus besonders sensibel sein. Trendanalysen oder Gruppenvergleiche scheinen robuster zu sein - allerdings nur dann, wenn das Editionsverfahren konstant bleibt. Insofern ist auch die Konsistenz der Editionsprozesse über verschiedene Teilstudien hinweg zu berücksichtigen. Ideal wäre eine parallele Konzeption und Entwicklung von Editions- und Auswertungsprogramm - dies scheint allerdings in arbeitsteiligen Großprojekten nur schwer möglich.

Außerdem zeigt es sich bei der Edition einmal mehr, dass die statistischen Resultate der empirischen Sozialforschung nicht übergenau, ‚bis in die dritte Nachkommastelle', inhaltlich interpretiert werden können. Neben der sozialen Realität gibt es auch im gesamten Forschungsprozess eine ganze Reihe von Faktoren, die die Ergebnisse beeinflussen.

Korrespondenzadresse

Dr. Steffen Hillmert
Max-Planck-Institut für Bildungsforschung
Forschungsbereich Bildung, Arbeit und gesellschaftliche Entwicklung
Lentzeallee 94
D-14195 Berlin
E-mail: hillmert@mpib-berlin.mpg.de

Literatur

Beck, U./Beck-Gernsheim, E., 1994: Individualisierung in modernen Gesellschaften - Perspektiven und Kontroversen einer subjektorientierten Soziologie. S. 10-39 in: dies. (Hrsg.): Riskante Freiheiten. Individualisierung in modernen Gesellschaften. Frankfurt/M.: Suhrkamp.

Brückner, H., 1995: Surveys don't lie, people do? An analysis of data quality in a retrospective life course study. Materialien aus der Bildungsforschung Nr. 50. Berlin: Max-Planck-Institut für Bildungsforschung.

Brückner, E./Hoffmeyer-Zlotnik, J./Tölke, A., 1983: Die Datenedition als notwendige Ergänzung der Datenerhebung bei retrospektiven Langzeitstudien. ZUMA-Nachrichten 13: 73-83.

Brückner, E./Mayer, K. U., 1998: Collecting life history data: Experiences from the German life history study. S. 152-181 in: J. Z. Giele/G. H. Elder Jr. (Hrsg.): Methods of life course research: Qualitative and quantitative approaches. Thousand Oaks: Sage.

Corsten, M./Hillmert, S., 2001: Qualifikation, Berufseinstieg und Arbeitsmarktverhalten unter Bedingungen erhöhter Konkurrenz. Was prägt Bildungs- und Erwerbsverläufe in den achtziger und neunziger Jahren? Arbeitspapier Nr. 1 des Projektes Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland. Berlin: Max-Planck-Institut für Bildungsforschung.

Hillmert, S./Kröhnert, S., 2001: Vergleich der Lebensverlaufsstudie 64/71-West mit Erhebungen des Mikrozensus auf Basis ausgewählter Randverteilungen. Technischer Arbeitsbericht. Berlin: Max-Planck-Institut für Bildungsforschung.

Infas, 1999: Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland. Methodenbericht zur Hauptstudie. Bonn: Institut für angewandte Sozialwissenschaft.

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

Schnell, R./Hill, P. B./Esser, E., 1995 (5. Aufl.): Methoden der empirischen Sozialforschung. München: Oldenbourg.

Spengemann, P./Hillmert, S., 2002: Editionshandbuch zur Studie Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland (unter Mitarbeit von H. Engelhardt, A. Mertens und H. Trappe; Stand 9/2002). Berlin: Max-Planck-Institut für Bildungsforschung.

Tölke, A., 1982: Möglichkeiten und Grenzen einer Edition bei retrospektiven Verlaufsdaten: dargestellt an der Lebensverlaufsbefragung des Sfb 3. Arbeitspapier Nr. 90. Frankfurt/M.: Universität.

Anhang

Tabelle 2: Veränderungen im Zuge der Edition: Variable *Ausbildungsabschluss* im Modul Ausbildungsgeschichte (ausgewählte Kategorien); Angaben in Prozent

	Vor Edition	Nach Edition
Ausbildungsabschluss (Alle Episoden)		
Praktikum/Volontariat	1,7	10,3
Gewerbliche Lehre	31,6	31,8
Kaufmännische Lehre	22,6	22,6
Berufsfachschulabschluss	3,0	6,5
Meister/Techniker	2,1	2,9
Fachschulabschluss	2,6	2,7**
Fachhochschulabschluss	3,3	4,4
Hochschulabschluss (Diplom)	5,9	3,6
Hochschulabschluss (1. Staatsexamen)	3,3*	2,1
Anderer Abschluss	21,0	0,2
Sonstige Kategorien	2,9 *(4 Kat.)*	12,9 *(15 Kat.)*
N (Episoden mit Abschluss) = 100%	*3374*	*3699*
Ausbildungsabschluss (Erste Episode)		
Praktikum/Volontariat	1,4	9,8
Gewerbliche Lehre	39,7	39,8
Kaufmännische Lehre	26,2	26,4
Berufsfachschulabschluss	2,9	5,3
Meister/Techniker	0,3	0,0
Fachschulabschluss	1,5	1,2**
Fachhochschulabschluss	1,8	2,4
Hochschulabschluss (Diplom)	4,8	3,4
Hochschulabschluss (1. Staatsexamen)	2,6*	2,3
Anderer Abschluss	15,9	0,1
Sonstige Kategorien	2,9 *(4 Kat.)*	9,3 *(15 Kat.)*
N (Personen mit Abschluss in erster Episode) = 100%	*2277*	*2341*
Veränderung der Episodenanzahl auf der Individualebene		
Reduktion um 2 und mehr	-	0,8
Reduktion um 1	-	6,1
Konstant	-	78,4
Zunahme um 1	-	11,3
Zunahme um 2 und mehr	-	3,3
N (Personen) = 100%		*2904*

* alle Staatsexamina; ** ohne Meister/Techniker.
Westdeutsche Lebensverlaufsstudie 64/71, eigene Berechnungen.

Tabelle 3: Veränderungen im Zuge der Edition: Variable Differenzierte *berufliche Stellung* im Modul Erwerbsgeschichte (ausgewählte Kategorien); Angaben in Prozent

	Vor Edition	Nach Edition
Diff. berufliche Stellung (Alle Episoden)		
Angestellte mit einfacher Tätigkeit	8,2	7,4
Angestellte mit qualifizierter Tätigkeit	37,4	35,9
Angestellte mit hochqualifizierter Tätigkeit	8,7	8,8
Angestellte mit höherer Leitungsfunktion	1,6	0,9
Ungelernte Arbeiter	5,6	5,2
Angelernte Arbeiter	9,2	8,4
Facharbeiter	19,4	19,8
Vorarbeiter	1,4	1,4
Sonstige Kategorien	8,5 *(19 Kat.)*	12,2 *(28 Kat.)*
N (Episoden mit benennbarer beruflicher Stellung) = 100%	*6826*	*7359*
Diff. berufliche Stellung (Erste Episode)		
Angestellte mit einfacher Tätigkeit	8,3	7,3
Angestellte mit qualifizierter Tätigkeit	40,4	39,5
Angestellte mit hochqualifizierter Tätigkeit	6,5	6,4
Angestellte mit höherer Leitungsfunktion	0,6	0,2
Ungelernte Arbeiter	6,0	5,6
Angelernte Arbeiter	7,3	6,3
Facharbeiter	24,5	24,6
Vorarbeiter	0,6	0,6
Sonstige Kategorien	5,8 *(19 Kat.)*	9,5 *(28 Kat.)*
N (Personen mit benennbarer beruflicher Stellung in erster Episode) = 100%	*2628*	*2741*
Veränderung der Episodenanzahl auf der Individualebene		
Reduktion um 2 und mehr	-	1,7
Reduktion um 1	-	5,4
Konstant	-	76,2
Zunahme um 1	-	12,7
Zunahme um 2 und mehr	-	3,8
N (Personen) = 100%		*2904*

Westdeutsche Lebensverlaufsstudie 64/71, eigene Berechnungen.

Tabelle 4: Veränderungen im Zuge der Edition: Ausbildungsverteilung in bestimmten beruflichen Stellungen (ausgewählte Kategorien); Angaben in Prozent

	Vor Edition			Nach Edition		
Ausbildungsabschluss*:	Ohne	Beruf	Studium	Ohne	Beruf	Studium
Berufliche Stellung (Alle Erwerbsepisoden)						
Angestellte mit einfacher Tätigkeit	45,4	51,9	2,6	33,3	64,3	2,4
Angestellte mit qualifizierter Tätigkeit	25,4	69,1	5,5	6,2	88,9	5,0
Angestellte mit hochqualifizierter Tätigkeit	14,0	38,4	47,6	5,9	45,0	49,1
Angestellte mit höherer Leitungsfunktion	17,1	60,4	22,5	4,5	68,2	27,3
Ungelernte Arbeiter	59,2	37,4	3,4	56,3	41,4	2,3
Angelernte Arbeiter	42,3	56,9	0,8	37,5	61,7	0,8
Facharbeiter	14,6	85,2	0,2	4,9	94,9	0,2
Vorarbeiter	11,5	88,5	0,0	7,5	92,5	0,0
...						
N (Episoden mit benennbarer beruflicher Stellung)			*6934*			*7276*
Berufliche Stellung (Erste Erwerbsepisode)						
Angestellte mit einfacher Tätigkeit	52,1	44,7	3,2	41,9	55,2	3,0
Angestellte mit qualifizierter Tätigkeit	25,3	69,0	5,7	6,6	88,4	5,0
Angestellte mit hochqualifizierter Tätigkeit	11,8	14,1	74,1	5,2	21,6	73,2
Angestellte mit höherer Leitungsfunktion	11,8	58,8	29,4	0,0	57,1	42,9
Ungelernte Arbeiter	73,9	22,3	3,8	71,6	26,4	2,0
Angelernte Arbeiter	51,6	47,4	1,0	45,9	51,9	2,2
Facharbeiter	12,8	87,1	0,2	4,4	95,6	0,0
Vorarbeiter	20,0	80,0	0,0	5,9	94,1	0,0
...						
N (Personen mit benennbarer beruflicher Stellung in erster Episode)			*2624*			*2595*

* Ohne: ohne Ausbildungsabschluss; Beruf: abgeschlossene Berufsausbildung; Studium: abgeschlossenes (Fach-)Hochschulstudium; die Angaben ergänzen sich für jede berufliche Stellung zu 100 Prozent.

Westdeutsche Lebensverlaufsstudie 64/71, eigene Berechnungen.

SCIENTIFIC-USE-FILE DES MIKROZENSUS 1998 FÜR DIE WISSENSCHAFT ZUGÄNGLICH

Beim Statistischen Bundesamt können ab Januar 2003 die faktisch anonymisierten Einzeldaten des Mikrozensus 1998 für Forschungszwecke bestellt werden. Wie für die bisher verfügbaren Mikrozensus-Scientific-Use-Files erfolgt die Lieferung der Daten zu Sonderkonditionen (Bereitstellungsgebühr von € 65,--), die durch eine Vereinbarung zwischen dem Statistischen Bundesamt und dem Bundesministerium für Bildung und Forschung (BMBF) ermöglicht wurden.

Die Daten der amtlichen Statistik sind für die empirischen Sozial- und Wirtschaftswissenschaften eine der wichtigsten Informationsquellen im Rahmen der Analyse wirtschafts- und gesellschaftspolitischer Zusammenhänge. Dies gilt insbesondere für den Mikrozensus, der als größte Mehrthemenumfrage der Bundesrepublik Deutschland vielfältige Informationen über die wirtschaftliche und soziale Lage der Bevölkerung enthält.

Das jährliche Grundprogramm umfaßt folgende Themenbereiche: sozio-demographische Angaben, Staatsangehörigkeit, Familien- und Haushaltszusammenhänge, Beziehung zum Lebenspartner (nichteheliche Lebensgemeinschaft), Haupt- und Nebenwohnung, Erwerbsbeteiligung und –tätigkeit, Schulbesuch, Ausbildungsabschluß, Quellen des Lebensunterhalts, Einkommen, Renten- und Pflegeversicherung. Darüber hinaus werden in jährlichen Ergänzungsprogrammen mit einem Auswahlsatz von 0,5 Prozent weitere Themen erfragt; 1998 umfasst dies ergänzende Angaben zur Erwerbstätigkeit und zur früheren Erwerbstätigkeit, Retrospektivangaben zur Situation ein Jahr vor der Erhebung (Erwerbstätigkeit, Stellung im Beruf, Wirtschaftszweig, Wohnsitz), Arbeitsstätte, Arbeitszeiten, Aus-, Fort- und Weiterbildung, Umschulung und Pflegebedürftigkeit. Im Rahmen der vierjährigen Zusatzprogramme enthält der Mikrozensus 1998 Angaben zur Wohnsituation (Art und Größe der Gebäude mit Wohnraum, Nutzung der Wohneinheit als Eigentümer, Haupt- oder Untermieter, Fläche und Baualter, Einzugsjahr, Heizungs- und Energieart, Warmwasserversorgung, Miete).

Nach dem Bundesstatistikgesetz von 1987 dürfen für die Durchführung wissenschaftlicher Vorhaben vom Statistischen Bundesamt und den Statistischen Ämtern der Länder

Einzelangaben an Hochschulen oder sonstige Einrichtungen in Deutschland mit der Aufgabe unabhängiger wissenschaftlicher Forschung übermittelt werden, wenn sie den Befragten nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft („faktische Anonymität“) zugeordnet werden können. Diese Daten müssen zweckgebunden verwendet werden und sind zu löschen, sobald das wissenschaftliche Vorhaben beendet ist. Für die Weitergabe einer, dem konkreten Forschungsvorhaben entsprechenden Variablenauswahl steht das Scientific-Use-File als eine faktisch anonymisierte 70%-Substichprobe des Mikrozensus 1998 zur Verfügung, das Einzelangaben über rund 510.000 Personen bzw. rund 230.000 Haushalte enthält.

Die als Scientific-Use-Files zur Verfügung stehenden Mikrozensus-Daten können beim Statistischen Bundesamt bestellt werden. Nähere Informationen dazu unter:

Statistisches Bundesamt
Zweigstelle Bonn
Gruppe IX B – Mikrozensus
Holger Breiholz
Postfach 17 03 77
53029 Bonn
Tel.: 01888/643-8711, Fax: -8692
E-mail: mikrozensus@statistik-bund.de

Darüberhinaus können sich interessierte Wissenschaftler an Ansprechpartner im Zentrum für Umfragen, Methoden und Analysen (ZUMA) in Mannheim wenden . Die Abteilung Mikrodaten bietet zu den Scientific-Use-Files im World Wide Web (http://www.gesis.org/Dauerbeobachtung/Mikrodaten/Daten/Abteilungsdaten/Mikrozensen/mz_daten_index.htm) u.a. folgende Informationen an: Datendokumentation, Fragebogen, rechtliche Voraussetzungen der Datenlieferung und Bestellmöglichkeiten.

ZUMA
Abteilung Mikrodaten
Postfach 12 21 55
68072 Mannheim
Tel.:0621/1246-265, Fax: -100
E-mail: mikrodaten@zuma-mannheim.de
http://www.gesis.org/Dauerbeobachtung/Mikrodaten/

LIS-SOMMERWORKSHOP
INTERNATIONALE VERGLEICHE DER EINKOMMENSUNGLEICHHEIT

21. – 26. JULI 2003, LUXEMBOURG

Die Luxembourg Income Study (LIS) ist eine internationale Einrichtung, die sich auf die Mikrodatenanalyse von Einkommensungleichheiten spezialisiert hat. Z. Zt sind dort in einer Datenbank über 100 große Mikrodatensätze gesammelt, die „harmonisiert", d.h. bezüglich mehrerer Indikatoren der Einkommenssituation und der ökonomischen Wohlfahrt für vergleichende Analysen erarbeitet worden sind. Die LIS-Datenbank umfasst derzeit Daten aus 26 Ländern: Australien, Österreich, Belgien, Kanada, Mexico, Dänemark, Finnland, Frankreich, Deutschland, Ungarn, Irland, Israel, Italien, Luxembourg, Niederlande, Norwegen, Polen, Russland, Tschechische Republik, Slowakische Republik, Spanien, Schweden, Schweiz, Taiwan, Großbritannien und USA.

ZUMA ist an der Finanzierung von LIS beteiligt und ermöglicht damit deutschen Forschern den freien Zugang zu den LIS-Daten.

Der LIS Sommer Workshop dauert eine Woche und ist darauf zugeschnitten, junge Forscher aus den Sozialwissenschaften (vor und nach der Graduierung) in die vergleichende Forschung von Einkommensverteilung und Sozialpolitik unter Verwendung der LIS-Datenbank einzuführen.

Die Unterrichtssprache ist Englisch. Der Sommerworkshop besteht aus einer Reihe von Vorlesungen und Tutorien, in denen die Teilnehmer mit ihren Forschungsthemen bei der Nutzung der LIS-Datenbank vom LIS Team (einschließlich Tim Smeeding – LIS Projektleiter, Lee Rainwater – LIS Forschungsleiter und John Coder LIS Technischer Leiter) und von anderen erfahrenen LIS-Nutzern beraten und unterstützt werden.

Der Workshop findet im Institut d'Europe statt, das sich im historischen Chateau de Munsbach in Luxembourg befindet. Der Workshop beginnt am 21. Juli (Anreise am Abend des 20. Juli) und endet am 26. Juli. Die Teilnahmegebühr beträgt 800 US $ bei Unterbringung in Zwei- oder Drei-Bett-Zimmern (Studentenwohnheim) und 1400 US $

bei Unterbringung im Hotel. Die Gebühr schließt Unterrichtsmaterial, Unterbringung und Verpflegung mit ein. Reisekosten sind vom Teilnehmer selbst zu tragen.

Formulare für die Bewerbung um einen Platz können von der LIS homepage www.lisproject.org/workshop.htm heruntergeladen werden und müssen ausgefüllt bis zum 15. April 2003 eingesandt werden. Es ist zu beachten, dass das Kontingent an Plätzen beschränkt ist.

Nähere Informationen über den Workshop kann man erhalten von:

Caroline de Tombeur
LIS Administrative Assistant
17, rue des Pommiers
L-2343 Luxembourg Stadt
Luxembourg
Tel: 00352 26003020
caroline@lisproject.org

Kati Foley
LIS Administrative Assistant
426 Eggers Hall
Syracuse University
Syracuse, NY 13244-1020 USA
Tel: 001 315 443-4526
lisaa@maxwell.syr.edu

BUCHBESPRECHUNG

Um es vorneweg zu sagen: Mit seinem Buch „Das Feldexperiment in der Umfrageforschung“ legt ein sehr belesener Autor Thomas Petersen ein sehr lesenswertes Buch vor. Geschrieben in der Tradition von „Alle, nicht jeder. Einführung in die Methoden der Demoskopie“ (Noelle-Neumann & Petersen 1998, 2. Auflage) – also leicht, aber nicht seicht, amüsant, aber nicht albern, lehrreich, aber nicht lehrmeisterhaft – lesen wir ein Buch, von dem wir am Ende sagen werden, das Lesen hat Spaß gemacht, und wir haben – dennoch – einiges gelernt. Hätte es der Autor dabei belassen, das Feldexperiment als ein Verfahren innerhalb der Umfrageforschung und seinen Nutzen für selbige darzustellen, hätte das Buch keinen Ansatz für (negative) Kritik geboten. Warum aber muss Petersen das Feldexperiment als Beleg dafür missbrauchen, dass die empirische Sozialforschung noch immer eine unterentwickelte Wissenschaft sei? Mehr dazu gleich. Sehen wir uns „Das Feldexperiment in der Umfrageforschung“ etwas näher an.

THOMAS PETERSEN
Das Feldexperiment in der Umfrageforschung
Frankfurt: Campus Verlag, 2002
ISBN 3-593-37014-X, 283 Seiten, € 39

Das Buch beschäftigt sich mit dem kontrollierten Feldexperiment, einem Verfahren, das, so Petersen, „vor allem in der deutschen empirischen Sozialforschung bisher wenig angewandt wird“ (S. 7). Der Schwerpunkt liegt dabei auf der „gegabelten Befragung“, dem „split ballot“. Bei dem Verfahren, bei dem insidern sofort das Institut für Demoskopie in Allensbach einfällt, wird die Stichprobe einer Umfrage in zwei oder mehr gleich große Teilstichproben unterteilt, die gleichzeitig und unter gleichen Bedingungen mit einem Fragebogen befragt werden, der sich für die zwei oder mehr Teilstichproben nur in den zu testenden Aspekten unterscheidet, ansonsten aber völlig identisch ist. Die Variation des Fragebogens stellt den experimentellen Stimulus dar. Wie noch gezeigt werden wird, können mit split ballot–Verfahren aber auch andere als Fragebogen-Variationen getestet werden.

Das Buch besteht aus drei wesentlichen Teilen. In den ersten drei Kapiteln wird die Bedeutung des Experimentes für die Entwicklung einer Wissenschaft im Allgemeinen und die Bedeutung des Feldexperimentes für die Entwicklung der empirischen Sozialforschung im Besonderen beschrieben; Kapitel 4 und 5 befassen sich mit theoretischen Überlegungen zur gegabelten Befragung und mit ihren Einsatzmöglichkeiten in der angewandten Sozialforschung; in den Kapiteln 6 bis 9 wird gezeigt, wie sich mit split ballot-Verfahren „Grundstrukturen des Befragtenverhaltens im sozialwissenschaftlichen Interview erforschen lassen, deren Kenntnis wiederum die Voraussetzung für die Entwicklung präziserer Fragebogenmethoden ist" (S. 8).

In der **Einleitung** gibt Petersen zunächst einen historischen Rückblick auf die Entwicklung des Experiments. In dieser zwar kurzen, aber anschaulichen Darstellung – interessant und lesenswert für Menschen, die an Wissenschaftsgeschichte interessiert sind – verfolgt Petersen den Weg des Experimentes von Roger Bacon (1210-1294), dessen Werk *Opus Maius* nach Petersen „in gewisser Hinsicht" als der „Beginn der Entwicklung zur Experimentalwissenschaft" verstanden werden kann (S. 14), über Francis Bacon (1561-1626), der „vehement die Einführung des *experimentum* in die Wissenschaft" gefordert habe, dessen Experiment-Begriff aber „unscharf" geblieben sei (S. 15) und John Stuart Mill (1806 – 1873), der nach Petersen zum ersten Male die „modernen Konzepte des Experiments" dargestellt und darauf hingewiesen habe, „daß mit der experimentellen Methode Kausalzusammenhänge aufgedeckt werden können" (S. 16). In dem Abschnitt über das Experiment in den Naturwissenschaften lesen wir von Entwicklungen und Personen, die man in sozialwissenschaftlichen Abhandlungen üblicherweise nicht findet: In den Naturwissenschaften würden in systematischer Weise Experimente ab dem 17. Jahrhundert durchgeführt, auch wenn es vorher schon vereinzelte Experimente oder empirische Arbeiten gegeben habe. In der Astronomie beginne die Zuwendung zur Empirie mit Kopernikus' (1543) Arbeit „Über die Umläufe der Himmelskreise", die, auch wenn es sich nicht um eine empirische Arbeit handle, „deutliche empirische Züge" trägt (S. 21). In den folgenden Jahrzehnten komme es zu einer immer stärkeren Zuwendung zur empirischen Beobachtung (Brahe, Kepler). Galilei schließlich gehe „den Schritt von der passiven Beobachtung zum aktiven Experiment, bei dem der Forscher absichtlich und geplant in die Realität eingreift und den Zustand, den er beobachten will, selbst herbeiführt" (S. 23). Zu nennen seien auch Stevin, Galilei, Gilbert, Newton in der Physik. Auch in der Medizin führe der Weg von der Theorie zur Empirie und von der Empirie zum Experiment: Paracelsus, Vasalius hätten schon empirisch, aber noch nicht experimentell gearbeitet. Die experimentelle Methode wäre von Santoro (1561 – 1636) in die Medizin eingeführt worden. Damit die Einleitung nicht nur deskriptiven Charakter hat, zieht Petersen aus dem Gesagten das für das gesamte Buch entscheidende Fazit: „In allen ... (naturwissenschaftli-

chen; rp) Wissenschaftsbereichen entwickelte sich zuerst das Bewußtsein der Notwendigkeit empirischer Beobachtung. Dann erst, teilweise mit erheblichem Zeitabstand, setzte sich das Experiment als Forschungsmethode durch. Und praktisch unmittelbar nach der Durchsetzung der experimentellen Methode beschleunigte sich der Fortschritt in der jeweiligen Wissenschaft beträchtlich." (S. 28). Petersen geht es jetzt darum, aufzuzeigen, „daß dieser Zweischritt in der Entwicklung der Wissenschaft auch in den Sozialwissenschaften beobachtet werden kann und daß sich verschiedene Disziplinen der Sozialwissenschaften heute vermutlich in unterschiedlichen Entwicklungsstadien befinden". (S. 29).

Dennoch: In weiten Teilen der Sozialwissenschaften stehe die Entwicklung des Experimentes noch immer am Anfang. Warum das so ist, beleuchtet Petersen im **zweiten Kapitel** über das **Experiment in der Sozialforschung**. Da seien zum einen moralische Bedenken: Ist es gerechtfertigt, Experimente an Menschen vorzunehmen? Extrembeispiele wie das Milgram-Experiment brächten den ganzen experimentellen Ansatz in Verruf. Da seien zum zweiten konzeptionelle und logische Probleme, die sich aus den Eigenschaften des Untersuchungsgegenstandes ergäben: Einerseits bewirke die Komplexität und Vielfalt sozialer Erscheinungen eine Schwierigkeit beim Umdenken vom Individual- zum Merkmalsbereich. Andererseits könne die Logik des naturwissenschaftlichen Experimentes nicht in die Sozialwissenschaften übernommen werden, weil Störfaktoren auftreten können und man nicht sicher sein kann, dass ein experimentelles Ergebnis wirklich aufgrund der experimentellen Bedingung entsteht. Und schließlich: Man könne nicht – wie in den Naturwissenschaften – von einem „Normalzustand" (Wasser = flüssig) der Untersuchungseinheiten ausgehen. Vorher-Nachher-Messungen hätten in den Sozialwissenschaften deshalb keinen Sinn. Statt dessen sieht Petersen in der Parallelgruppenuntersuchung die „Lösung des Problems" (S. 40): „Erst als dieser Schritt vom Vorher-Nachher-Modell zum Parallelgruppenmodell vollzogen war, konnte das Experiment auch in den Sozialwissenschaften nutzbar gemacht werden." (S. 41). Aufgrund der Schwierigkeiten, das Experiment in die Sozialwissenschaften zu übernehmen, sei es in unterschiedlichen sozialwissenschaftlichen Disziplinen unterschiedlich vertreten; als Schwerpunkte nennt Petersen die physiologische Psychologie und die Sozialpsychologie, ohne allerdings zu erklären, warum etwa in der Sozialpsychologie die gerade genannten Hinderungsgründe für das Experiment weniger wirksam sein sollten als in der empirischen Sozialforschung, in der sich die experimentelle Methode bisher nur in Ansätzen habe durchsetzen können, und das noch dazu nur in den Vereinigten Staaten. Dort sei im Jahre 1940 zum ersten Male über split ballot-Experimente berichtet worden (Roper; Benson; Cantril). 1944 erscheint Cantrils Buch „Gauging Public Opinion" als erstes Buch über split ballot in Bevölkerungsumfragen, 1951 dann Payne's „The Art of Asking Questions". Danach herrscht über 30 Jahre Ruhe, bis Elisabeth Noelle-Neumann 1970 ihren Beitrag „Wanted: Rules for

Wording Structured Questionnaires“ publiziert. Trotzdem gibt es heute in den USA „keine Institution, die kontinuierlich Feldexperimente im Rahmen von Repräsentativumfragen durchführt“ (S. 45). Das gleiche gilt für Deutschland, wenn man vom Institut für Demoskopie absieht: „das Allensbacher Institut blieb die einzige Institution, die solche Verfahren kontinuierlich, regelmäßig und systematisch anwandte“ (S. 46). Dennoch ist Petersen zumindest für die USA optimistisch: „Während also in der deutschen empirischen Sozialforschung die experimentelle Methode nach wir vor nur in außerordentlich geringem Umfang genutzt wird, gewinnt sie in den Vereinigten Staaten ganz allmählich an Bedeutung. Es hat den Anschein, als vollziehe dort die Sozialforschung langsam den zweiten Schritt in der Wissenschaftsentwicklung, den Schritt von der Empirie zum Experiment, der in der deutschen Umfrageforschung noch bevorsteht.“ (S. 49). Die Begründung, warum das so ist, wäre zu diskutieren: Traditionellerweise habe sich die deutsche Sozialforschung zu Beginn weniger durch naturwissenschaftlich oder psychologisch orientiertes Denken ausgezeichnet als durch sozialreformerische Bestrebungen, was sich noch heute in einem Schwerpunkt der Soziologie im Bereich sozialer Schichtung und damit verbundener Benachteiligung bestimmter Bevölkerungsgruppen nachzeichnen lasse; die amerikanische Sozialforschung hingegen habe schon sehr früh den Kontakt zur Sozialpsychologie gefunden und dadurch eine Tradition des public opinion research entwickelt.

Nach diesem Ausflug in die Wissenschaftsgeschichte und nach eher allgemeinen Überlegungen zum Experiment in der Sozialforschung wendet sich Petersen mit **Kapitel 3** dem **Feldexperiment** zu. Von diesem sagt er: „Feldexperimente sind wegen der häufig größeren Unschärfe und geringeren Eindeutigkeit ihrer Ergebnisse zum Aufspüren detaillierter Zusammenhänge oft weniger gut geeignet als Laborexperimente. Doch sie können dazu dienen, Laborexperimente zu validieren, indem sie feststellen, wie sehr ein im Labor entdeckter Faktor tatsächlich wirksam ist.“ (S. 66). Und: ein weiterer Vorteil des Feldexperimentes auf der Basis repräsentativer Befragungen sei, daß sie ein „eigenes Entdeckungspotential ... in allen Bereichen tatsächlichen sozialen Handelns...“ hätten (ebenda). Als Beispiele dafür nennt Petersen „soziale Experimente“ (z.B. das Kabelpilotprojekt Ludwigshafen), Quasi-Experimente (bei denen erst nach Eintreten eines Ereignisses als experimentellem Faktor die Personen der Test- bzw. der Kontrollgruppe zugeordnet werden) und Extremgruppenanalysen (bei denen unterschiedliche Stichproben gebildet werden, die sich nur in einem Punkt, nämlich dem interessierenden, unterscheiden; die Variation bestehe darin, daß die Versuchspersonen nach dem Gegenstand des Forschungsinteresses sortiert werden).

In **Kapitel 4** wendet sich Petersen nun der **gegabelten Befragung** zu, der seiner Ansicht nach „wahrscheinlich variationsreichsten und wissenschaftlich ertragreichsten Form des Feldexperimentes in der Umfrageforschung“ (S. 81). Grundprinzip: „im Rahmen einer

standardisierten repräsentativen Umfrage wird die Stichprobe in zwei (oder mehr) jeweils gleich große Teilstichproben unterteilt. Jede dieser Teilstichproben ist gleichermaßen repräsentativ für die Grundgesamtheit. Die Personen jeder Teilstichprobe werden gleichzeitig und unter gleichen Bedingungen befragt, und zwar mit einem Fragebogen, der sich von Gruppe zu Gruppe nur in Details unterscheidet, wie zum Beispiel in der Formulierung einer bestimmten Frage, der Variation der Reihenfolge zweier Fragen, oder durch leicht unterschiedliche Bildvorlagen. Diese Variationen des Fragebogens bilden den experimentellen Stimulus, dem die Hälfte der Befragten ausgesetzt wird" (S. 81). Die Frage, ob es sich dabei tatsächlich um Experimente handle, werde kontrovers diskutiert; aus Petersen's Sicht aber wird sie eindeutig bejaht, erfüllten Feldexperimente doch die vier Bedingungen, die nach Schulz (1970) für ein Kontrollgruppenexperiment zutreffen müssten: Gleichheit der Experimental- und der Kontrollgruppe in allen Faktoren, Bildung der Gruppen vor Anbringung des experimentellen Faktors und zeitgleiches Anbringen des Faktors in den Gruppen, vollkommene Abschirmung der Kontrollgruppe vom experimentellen Faktor und Unterscheidung von Experimental- und Kontrollgruppe nur und nur durch den experimentellen Faktor: „Wenn also bei einem Split-Ballot-Experiment die Befragten der einen Teilstichprobe anders auf einen im Fragebogen untergebrachten Stimulus reagieren als die Befragten der anderen Teilstichprobe auf das an der entsprechenden Stelle des Vergleichsfragebogens angebrachte Signal, kann dieser Unterschied ... mit hoher und berechenbarer Wahrscheinlichkeit kausal auf die Variation des Stimulus im Fragebogen zurückgeführt werden. Der mögliche Einfluß einer unbekannten Drittvariablen läßt sich weitgehend ausschließen. Damit sind Split-Ballot-Experimente eindeutig echte, zum Kausalnachweis uneingeschränkt taugliche Experimente" (S. 85), wenngleich Petersen durchaus Begrenzungen des Einsatzes von split ballot-Experimenten bzw. Feldexperimenten insgesamt erkennt: enge Grenzen für die Komplexität der Stimuli, möglichst einfache Anforderungen an Interviewer und Befragungspersonen, der Interviewer ist kein Forscher wie im Laborexperiment, die Befragung findet bei den Befragungspersonen und nicht im Labor statt, umfangreiche Itembatterien sind im Feldexperiment nicht einsetzbar, mehr als ein oder zwei experimentelle Faktoren können nicht eingebaut werden.

In **Kapitel 5** geht Petersen der Frage nach, in welchen **Bereichen** split ballot-Verfahren **zum Einsatz** kommen können. Er nennt zunächst die **Verhinderung von Reihenfolge**- (z. B. durch Variation der Fragen- bzw. Itemreihenfolge) und **Kontexteffekten**; bei letzteren funktionieren die häufig zum Einsatz gebrachten Puffer-Fragen allerdings nicht immer: „So bleibt letztlich nur die Möglichkeit übrig, die Fragen, von denen man vermutet, daß sie Kontext-Effekte hervorrufen, jeweils nur einer Halbgruppe zu stellen..." (S. 104). Dass damit die unangenehme Nebenwirkung auftaucht, dass man die Fallzahl durch ein solches

Vorgehen halbiert, sieht Petersen durchaus kritisch, nimmt sie aber in Kauf. Als zweiten Einsatzbereich für split ballot-Verfahren nennt Petersen den **Test von Frageformulierungen**, für den er, wie in dem gesamten Buch überhaupt, viele Beispiele aus der Literatur und aus Allensbach-Umfragen anführt; vor allem in den frühen vierziger Jahren des 20. Jahrhunderts sei split ballot mit Test von Frageformulierungen gleichbedeutend gewesen (z. B. balanced vs. unbalanced Formulierungen, harte vs. weiche Formulierungen, Ja-Sage-Tendenz). Das Problem dabei sei, dass die Ergebnisse keinen Hinweis darauf lieferten, welche der getesteten Varianten die „bessere" sei, welche die tatsächliche Verteilung in der Bevölkerung wiedergebe. Und: Manchmal seien die erwarteten Effekte eben aufgetreten, manchmal dagegen nicht. Als dritten Anwendungsbereich nennt Petersen die **analytische Prüfung der Festigkeit eines Meinungsbildes in der Bevölkerung**, die man daran erkennen könne, wie groß die Effekte sind, die bei Fragebogenexperimenten zu dem betreffenden Thema auftreten. Gewagt erscheint, dass Petersen in diesem Zusammenhang nach Zitierung einer einzigen Studie von Hippler und Schwarz die Folgerung zieht: „Nach diesem Ergebnis kann man es bis auf weiteres als hinreichend belegt ansehen, daß ein Zusammenhang zwischen dem Auftreten von Fragebogeneffekten und der Entschiedenheit einer Bevölkerungsmeinung existiert". (S. 119). Nachdem die bisher beschriebenen drei Einsatzfelder als Belege dafür gelten können, wie die split ballot-Verfahren „zur Sicherung der Validität oder zur Beurteilung der Aussagekraft von Ergebnissen einzelner Fragen eingesetzt wird" (S. 127), folgen jetzt „Beispiele für Untersuchungskonzepte, „in denen die gegabelte Befragung die zentrale, zur Beantwortung der jeweiligen Forschungsfrage notwendige Methode ist" (S. 127). Da ist zum einen der **Test verschiedener Vorlagen und Argumente**, zum zweiten der **Einsatz von Feldexperimenten in Umfragen für die Rechtspraxis** (etwa zum Nachweis eines Kausalzusammenhangs zwischen Stimulus und Befragtenresponse) und zum dritten die Verknüpfung von Indikator-Fragen und Feldexperimenten als **Mittel zur Aufdeckung bis dahin unbekannter Zusammenhänge**, für die es noch keine theoretische Erklärung gibt. Um abzuklären, ob die Indikatorfrage tatsächlich die zugrundeliegende Dimension abbilde, könne man Indikatorfragen mit split ballot-Experimenten kombinieren. Dadurch ließen sich, so Petersen, „unbewußte Einstellungen der Befragten nicht nur zeigen, sondern nachweisen" (S. 143).

Mit **Kapitel 6** beginnt Petersen dann den dritten Teil des Buches, in dem gezeigt wird, wie sich mit der Methode der gegabelten Befragung **das Verhalten der Befragten im Interview analysieren** lässt: Unter welchen Bedingungen treten Kontext- und Reihenfolgeeffekte auf, und gibt es Muster im Auftreten dieser Effekte, die auf die kognitiven oder emotionalen Vorgänge bei den Befragten schließen lassen, welche die Beantwortung einer

Frage beeinflussen. Split ballot-Experimente seien geeignet zur Analyse des Befragtenverhaltens.

Kapitel 7 des Buches ist **Reihenfolgeeffekten** gewidmet, einem Forschungsbereich, dem Petersen eine unübersichtliche Erkenntnislage attestiert. Reihenfolgeeffekte ergäben sich „aus der Position einer Frage im Fragebogen ... , aus der Position einer Antwortkategorie innerhalb einer Frageformulierung oder aus der Zahl der vorgegebenen Antwortkategorien, und zwar unabhängig vom konkreten Inhalt einer Frage". (S. 165). Zum Thema Reihenfolgeeffekte sei die Literatur teilweise widersprüchlich, vor allem fehle die theoretische Grundlage zur Erklärung von Reihenfolgeeffekten. Die kognitive Psychologie leiste hier zwar einiges an guter Arbeit, sei zugleich aber auch umgekehrt verantwortlich für einen Teil der Unübersichtlichkeit der Befunde aufgrund einer „eher zu intensiven theoretischen Orientierung, nämlich der Neigung, statistisch signifikante Befunde voreilig inhaltlich zu interpretieren, zu verallgemeinern und theoretisch einzuordnen" (S. 166). Auch auf die Frage, ob in Dialogfragen mit primacy- oder recency-Effekten zu rechnen sei, gebe es keine allgemeingültigen Antworten, aber theoretische Erklärungsversuche der kognitiven Umfrageforscher. So habe z.B. Norbert Schwarz anhand von drei Experimenten aus Allensbach-Daten nachgewiesen, dass auch bei Dialogfragen ein recency-Effekt auftritt, wenn die Argumente vorgelesen werden und ein primacy-Effekt, wenn die Argumente schriftlich präsentiert werden. Eine meta-Analyse von 371 Datensätzen bestätige dies allerdings nur zur Hälfte. Bestätigt würde, dass beim Vorlesen das zweite Argument bessere Chancen habe, ausgewählt zu werden. Würden die Argumente dagegen auf Bildblättern vorgelegt, ergäbe sich im Gegensatz zu Schwarz' Erwartungen das gleiche Ergebnis, nämlich ein Überwiegen von recency-Effekten. Zudem seien in beiden Gruppen die recency-Effekte deutlich stärker als die primacy-Effekte. Und: Je mehr Worte die Argumente enthielten, umso stärker sei der gemessene recency-Effekt.

Kapitel 8 beschäftigt sich mit **Reaktionen auf Signale der formalen Fragengestaltung**. Da sei zunächst der **Nachfrage-Effekt** zu nennen: Die Befragungsperson habe ein Interesse daran, das Interview möglichst schnell und mit möglichst wenig Aufwand zu absolvieren und dabei eine angenehme Gesprächssituation zu haben. Dieses Interesse, verbunden mit der Neigung, nicht unbedingt die beste, sondern die am leichtesten zu erreichende Antwort zu finden, könne dazu führen, „daß aus Sicht des Befragten die Antwortmöglichkeit die optimale ist, die am wenigsten Anstrengung erfordert, wenn sie nicht fundamental den eigenen Überzeugungen widerspricht" (S. 194/195). Deshalb vermieden Befragte Antworten, die eine Nachfrage erwarten lassen. Als zweite Reaktion auf Signale der formalen Fragebogengestaltung nennt Petersen solche, „die sowohl von der numerischen Beschriftung als auch von der optischen Ausgestaltung einer Skalenvorlage ausgehen können" (S. 202). Bei gleicher logischer Struktur von Fragen- oder Antwortvorgaben sei es

möglich, dass Befragungspersonen die ihnen präsentierten Skalenvorgaben einschließlich der angegebenen numerischen Werte inhaltlich interpretierten und ihre Antworten daran ausrichten. Und: Personen neigten dazu, sich in der Mitte einer Skala einzuordnen. Diese Tendenz tritt auf und ist auch bekannt; Petersen beschränkt sich nun allerdings nicht auf kognitionspsychologische Erklärungen, sondern erweitert den Erklärungsansatz dahingehend, „daß möglicherweise ein Teil der Reaktionen auf die formale Fragengestaltung, unter ihnen die Tendenz zur Mitte bei Skalenvorlagen, nicht alleine auf rationale, kognitive, sondern auf emotionale Vorgänge zurückgeführt werden kann“ (S. 216). Dies wird mit dem Konzept der Isolationsfurcht erklärt, demzufolge sich Menschen auch bei Befragungen eher in der Mitte orientierten, um zu vermeiden, sich neben die Mehrheit zu stellen.

Kapitel 9 schließlich wendet sich den **inhaltsbedingten Ergebnisverzerrungen** zu. Dabei handelt es sich zunächst um Kontext-Effekte, also „die Reaktion auf zusätzliche, über den objektiven sachlichen Inhalt hinausgehende Signale, die der Befragte aus dem Frageinhalt entnimmt, und die sein Antwortverhalten im Interview beeinflussen, in aller Regel ohne daß dies vom Forscher beabsichtigt ist“ (S. 225). Zu den wichtigsten dieser Effekte zählten Kontext-Effekt als „Effekte, bei denen eine Frage im Fragebogen die Reaktion der Befragten auf eine nachfolgende ... Frage beeinflußt“ (S. 225/226), die durch vage Frageformulierungen oder Themen von geringer Bedeutung bewirkt werden könnten. Auch sie, so Petersen, seien „vermutlich teilweise auf kognitive, teilweise aber auch auf emotionale Reaktionen der Befragten zurückzuführen“ (S. 226). Die zwei Arten der Kontext-Effekte, nämlich Assimilations- und Kontrasteffekte, ließen sich zwar auch auf kognitive Erklärungen zurückführen – „Es kann kaum Zweifel daran bestehen, daß kognitive Vorgänge die Grundvoraussetzung dafür sind, daß Kontext-Effekte überhaupt auftreten“ (S. 230) – doch würden sie dadurch „nur zum Teil erklärt“ (S. 230). Auch hier spielten vielmehr, so Petersen, emotionale Aspekte eine Rolle. An einem Beispiel demonstriert Petersen, wie positive oder negative Stimmungen produziert werden, die dann wiederum auf die Beantwortung anderer Fragen grundsätzlich ausstrahlen.

Neben den Kontexteffekten wirke sich – so Petersen – aber auch die Medien- und Alltagssprache auf das Antwortverhalten aus. An dieser Stelle machen sich Zweifel an der Stichhaltigkeit der Argumentation von Petersen breit. Während z.B. soziale Wünschbarkeit häufig als Folge der Anwesenheit Dritter im Interview und als Reaktion auf bedrohliche Fragen verstanden wird, hängt sie nach Petersen von der Mediensprache bzw. der Alltagssprache einer Fragenformulierung ab, auch wenn das auf den ersten Blick mit sozialer Erwünschtheit gar nichts zu tun zu haben scheint: „Dem Konzept der ‚Mediensprache‘ liegt die Beobachtung zugrunde, daß sich auch umfangreiche, detaillierte Medienberichterstattung über öffentliche Streitthemen auf eine begrenzte Anzahl von Argumentationsstrukturen und Argumenten stützt, die in den Artikeln und Beiträgen zu dem jeweiligen

Thema immer wieder erwähnt werden, und zwar häufig mit gleicher, oder nur unwesentlich variierter Wortwahl. Dabei werden häufig zur Charakterisierung und zur positiven oder negativen Bewertung eines Gegenstandes oder Standpunktes Begriffe verwendet, die im alltäglichen Gespräch nicht oder zumindest seltener als in den Medien verwendet werden, die aber kurz und eingängig sind und sofort verstanden werden...“ (S. 242/244). Dies mag ja so sein. Aber dass und wie sich das auf die Ergebnisse von Umfragen auswirkt, bleibt unklar, insbesondere wenn im Beispiel auf Seite 236 der Begriff „schmutzige Wäsche waschen“ der Mediensprache zugeordnet wird – ohne Begründung und zumindest mit des naiven Lesers Vorstellung, dass ein solcher Begriff faktisch eher in der Alltagssprache anzusiedeln wäre. Auch das im Beispiel auf Seite 244 und 245 präsentierte Ergebnis muss uns eigentlich nicht verwundern: Wenn ich ein „Institut für Wirtschaftsforschung“ in einer Halbvariante als „umstritten“ und in der anderen ohne diesen Zusatz beschreibe und dann frage, „Was meinen Sie: Klingt das glaubwürdig, was das Institut da sagt, oder klingt das eher unglaubwürdig?“, muss ich mich nun wirklich nicht wundern, wenn die Befragten dem „umstrittenen“ Institut weniger Glaubwürdigkeit zubilligen. Da muss ich beileibe nicht Begriffe wie „Mediensprache“ oder „Soziale Konditionierung“ bemühen: der Begriff „umstritten“ ist nun mal kein Wort, „das man formal als eine sachliche, prinzipiell neutrale Zusatzinformation ansehen kann“, sondern „enthält also anscheinend eine Wertung, ein Signal, auf das die Befragten im Interview reagieren“ (S. 245) – aber ob uns diese Erkenntnis wirklich bisher verborgen war? Auch die Ausführungen Petersens zu zwei gegenläufigen Effektbedingungen auf Seite 251ff., die auf der Hypothese basieren, „daß der Reaktion der Befragten auf eine in Mediensprache formulierte Frage ein der klassischen Konditionierung ähnlicher assoziativer Lernvorgang zugrundeliegt“, aber das Verhalten im Interview noch nicht alleine erklären kann, ist nicht auf den ersten, auch auf den zweiten Blick nicht nachvollziehbar, auch nicht durch die Erweiterung des Erklärungsansatzes durch die schon genannte Isolationsfurcht: „Wenn Befragtenreaktionen auf Fragen in Mediensprache tatsächlich als Vermeidungsverhalten unter sozialem Druck angesehen werden können, dann wären sie damit nichts anderes als eine Form ‚sozial erwünschten‘ Verhaltens: Angesichts der über die Frageformulierung vermittelten Isolationsbedrohungen gibt der Befragte eine Antwort, ... mit der er gesellschaftliche Isolation vermeiden möchte, von der er also annimmt, daß sie – und nur sie – in der Gesellschaft akzeptiert, also sozial erwünscht ist.“ (S. 253/254). Man beachte die konjunktive Formulierung dieser Aussage – hat Petersen vielleicht doch gewisse Bedenken, dass er mit Konzepten wie Mediensprache, sozialer Konditionierung und Isolationsfurcht mit theoretischen Kanonen auf empirische Spatzen zielt? Dazu hätte Petersen nun wirklich keinen Grund gehabt, wenn er das split ballot und alles, was daraus folgt, als Technik beschrieben hätte, die in Umfragen durchaus – wie er selbst immer wieder empirisch belegen kann - zu verwertungsrelevanten Konsequenzen führen kann. Manche seiner über

die methodischen Stärken des split half hinausgehenden Schlussfolgerungen allerdings, insbesondere diejenigen aus Kapitel 9, wirken etwas konstruiert, und es hätte den positiven Gesamteindruck des Buches nicht geschmälert, wenn er darauf verzichtet hätte.

Auch die allgemeine Konsequenz, die Petersen aus seinen Überlegungen in **Kapitel 10** zieht, dass nämlich das Feldexperiment die „letzte Instanz" sei, könnte durchaus als übertrieben verstanden werden. Wenn Petersen versucht, die Bedeutung des split ballots über die schon lobend erwähnten Konsequenzen für die Durchführung und Auswertung von Umfragen überzubetonen, bewegt er sich auf unsicherem Terrain. Oder muss man wirklich unbedingt zustimmen, wenn Petersen – noch dazu am Vergleich von Quoten- und Zufallsstichprobe (zu diesem Thema ist alles gesagt, was es zu sagen gibt) – der Ansicht ist, die Tatsache, dass das einhellige Beharren der Sozialforschung auf der Zufallsstichprobe und die durch das umgekehrte Beharren des Instituts für Demoskopie auf der Quotenstichprobe und die damit verbundene „Isolation des Instituts in der Fachwelt" (S. 261) als Beleg dafür gilt, dass „nicht das Experiment, nicht die Empirie ... in diesem Methodenstreit die Oberhand behalten (hat), sondern die theoretische Erwägung" (S. 261). Oder muss man gar zustimmen, wenn Petersen darin den Beweis sieht dafür, dass die Sozialwissenschaften noch unterentwickelt seien: In dem Konflikt zwischen Quote und Random „in dem die Theorie gegen die Empirie stand, gab die Theorie den Ausschlag. In einer entwickelten Wissenschaft hätte die Empirie den Ausschlag gegeben" (S. 262). Und: „Das Feldexperiment ist die konsequente Fortsetzung der Hinwendung der Sozialwissenschaften zur Empirie. Es ist der notwendige zweite Schritt in der Wissenschaftsentwicklung. Ihm kommt damit für die Zukunft der empirischen Sozialforschung eine entscheidende Bedeutung zu" (S. 263).

Zum Abschluss noch einmal: Mit seinem Buch „Das Feldexperiment in der Umfrageforschung" legt Thomas Petersen ein sehr lesenswertes Buch vor, das man erfahrenen wie auch weniger erfahrenen Sozialforschern wirklich guten Gewissens ans Herz legen kann. Der Eindruck, dass bestimmte Argumentationen an manchen Stellen überstrapaziert worden sind, mindert dieses Urteil nicht. Trotz dessen oder auch ohne dies ist das Buch uneingeschränkt empfehlenswert.

ROLF PORST

PUBLIKATIONEN

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

ZUMA-Methodenbericht 2002/02
Cornelia Züll & Juliane Landmann
Computerunterstützte Inhaltsanalyse: Literaturbericht zu neueren Anwendungen

ZUMA-Methodenbericht 2002/03
Janet Harkness, Evi Scholz & Stefan Stadler
ISSP 1999 Germany. Social Inequality III. ZUMA Report on the German Study

ZUMA-Methodenbericht 2002/04
Janet Harkness, Evi Scholz & Stefan Stadler
ISSP 2000 Germany. Environment II. ZUMA Report on the German Study

ZUMA-Methodenbericht 2002/05
Michael Blohm & Ursula Lamas Perez
ALLBUS-Bibliografie (18. Fassung, Stand: Juli 2002)

(NON)RESPONSE BEI WEB-BEFRAGUNGEN

Web-Befragungen – über das World-Wide-Web realisierte Fragebogenuntersuchungen – erfreuen sich zur Bearbeitung der unterschiedlichsten Fragestellungen in der Psychologie sowie verwandter sozialwissenschaftlicher Disziplinen wachsender Beliebtheit. Wie auch bei klassischen selbstadministrierten Fragebogenuntersuchungen ist bei allen Web-Befragungen damit zu rechnen, dass von einem Teil der potenziellen Befragten entweder überhaupt keine Angaben vorliegen bzw. nicht alle der insgesamt gestellten Fragen beantwortet werden. Diese ‚Nonresponse'-Problematik bildet den Ausgangspunkt des Buchs und führt zur zentralen Fragestellung: Wie lässt sich die (teilweise) Nicht-Partizipation an Befragungen vorhersagen bzw. psychologisch sinnvoll erklären?

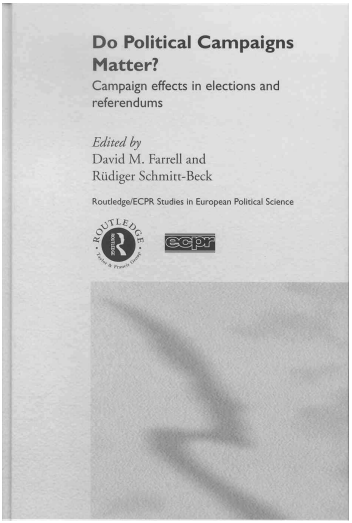
Während in der klassischen Umfrageforschung die meisten Arbeiten zum Nonresponse einer behavioristischen Logik folgen, die kaum Schlussfolgerungen über die der (Nicht-)Teilnahme an Befragungen zugrundeliegenden psychologischen Prozesse zulässt, liegt hier ein Versuch einer handlungstheoretischen Modellierung des Partizipationsverhaltens vor. Web-Befragungen bieten dabei einen besonders geeigneten Zugang zur Thematik, denn diese liefern weitaus zuverlässigere und detailliertere Einblicke in die

´Black-Box´ des Partizipationsprozesses als klassische selbstadministrierte Befragungsverfahren.

Im Rahmen von insgesamt vier empirischen Studien wurde untersucht, inwieweit unterschiedliche Formen der (Nicht-)Teilnahme an Web-Befragungen vorhersagbar und erklärbar sind. Die theoretische Grundlage bildete die ´Theorie des geplanten Verhaltens´ von Icek Ajzen, erweitert um einige dem Verhaltensbereich (Non)Response angemessene Determinanten und Moderatoren. Darüber hinaus wird ein Überblick zu den folgenden Themen gegeben: Verfahren der Teilnehmerrekrutierung und Stichprobenziehung für Web-basierte Befragungen, Formen des (Non)Response bei Web-Befragungen, Erklärungsansätze für (Non)Response, (Non)Response und handlungstheoretisch fundierte Persuasion, Zur Bedeutung der ermittelten Befunde bei der Auswahl von Verfahren zur Substitution bzw. Elimination fehlender Angaben.

(Non)Response bei Web-Befragungen
Auswahl, Erweiterung und empirische Prüfung eines handlungstheoretischen Modells zur Vorhersage und Erklärung des Partizipationsverhaltens bei Web-basierten Fragebogenuntersuchungen.
Michael Bosnjak
Aachen: Shaker-Verlag
ISBN 3-8322-0818-6, 299 Seiten, 49.80 €

DO POLITICAL CAMPAIGNS MATTER?



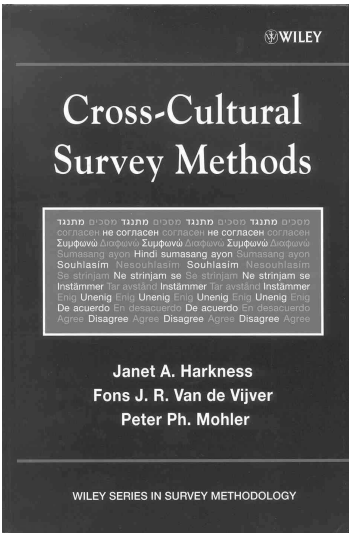
In recent decades, political actors of all sorts - Parties and candidates, governments and other political institutions, interest groups and social movements - have increasingly come to view political campaigning as an essential supplement to their engagement in the pro??cess of policy-making. By investing ever more efforts and resources into political campaigns, they seek to mobilize support among the mass public, to persuade citizens of their causes and to inform the citizenry about public policies and political procedures. So far as the practitioners are concerned, such campaigns matter a great deal. Each year, billions of dollars are spent on political campaigns. The sophisticated services of specialist agencies and campaign consultants are engaged; candidates are sent on television training courses; glossy literature, advertisements and campaign gimmicks are produced. While parties, candidates, interest groups, governments, media and - perhaps - voters all seem to be strongly convinced of the notion that campaigns do indeed matter, the collective views of the academic community can perhaps best be summarized as undecided.

This book, in bringing together some of the leading international scholars on electoral behaviour and communication studies, provides the first ever stock- take of the state of

this sub-discipline. The individual chapters present the Most recent studies on campaign effects in North America, Europe and Australia. As a whole, the book provides a cross-national assessment of the theme of political campaigns and their consequences.

Do Political Campaigns Matter?
Campaign Effects in Elections and Referendums
David M. Farrell & Rüdiger Schmitt-Beck (Editors)
Routledge/ECPR Studies in European Political Science 25
Publisher: Routledge 2002
ISBN 0415255937, Hardcover, 208 Pages, US $90.00

CROSS-CULTURAL SURVEY METHODS



Since all research is in a sense comparative, it is sometimes argued that methods and requirements remain the same, no matter what country or nationality is involved. The editors of this volume contend that this philosophy is an oversimplification. Although comparative survey research builds on and benefits from best practices established in monocultural research, cross-cultural research requires methodologies which address the central issues of equivalence of measurement and comparability of findings. The literature dealing with these issues has tended to be piecemeal, scattered across disciplines, and focused on specific questionnaires and the needs of single disciplines. This book provides the first systematic, interdisciplinary and hands-on treatment of the issues of greatest saliency for comparative survey research within and across countries.

Breaking new ground in its approach, Cross-Cultural Survey Methods describes how to recognize and deal with the major obstacles at each stage of researching, striving for equivalence and comparability. Topics include:

- Designing and crafting questionnaires for comparative subjects, including questionnaire translation
- Error and bias issues in cross-national surveys
- Techniques for analyzing bias and equivalence
- Statistical techniques for substantive analysis and the use of multidimensional scaling to analyze bias and research questions
- Important issues of preparing data for secondary analysis, such as data access, and documentation
- An introduction to meta-analysis in comparative survey research

Cross-Cultural Survey Methods
Janet Harkness, Fons van de Vijver & Peter Ph. Mohler (Eds.)
Wiley & Sons, Inc., Publication, 2002
ISBN 0-471-38526-3, Hardcover, 432 Pages, US $89.95

ZUMA-VERANSTALTUNGEN 2003

Workshop: Mitarbeiterbefragungen	18.–19. März
Workshop: Sozio-demographische Variablen in der Umfrageforschung	25.–26. März
Workshop: Einführung in die Clusteranalyse	01.–02. April
Workshop: Einführung in Datenmanagement und Datenrecodierung	16.–17. Juni
Workshop: Evaluation von Studiengängen – Methodische Aspekte	24.–25. Juni
Workshop: Einführung in die Analyse linearer Strukturgleichungsmodelle mit LISREL 8.52	26.–27. Juni
Workshop: Hermeneutic Procedures to Improve Questionnaire Research	02.–03. Juli
Workshop: ‚Political Polls': Politische Umfragen in den Massenmedien	10.–11. Juli
Workshop: Survey Design for the Multinational Context	04.–05. September
Workshop: Simulation for the Social Sciences	08.-12. September
Workshop: Analyse linearer Strukturgleichungsmodelle mit LISREL 8.52 für Fortgeschrittene	25.–26. September
Workshop: ADM-Stichprobensysteme in der Umfragepraxis: CATI und face-to-face	30. Sept.–1. Oktober
3. Nutzerkonferenz: Forschung mit dem Mikrozensus	9.–10. Oktober

Hinweise zu weiteren Veranstaltungen der GESIS finden Sie unter:
www.gesis.org/veranstaltungen/

* * *

Mitarbeiterbefragungen

18. UND 19. MÄRZ 2003

Mitarbeiterbefragungen (MABs) gehören zu den am häufigsten eingesetzten Umfragen. Sie bilden eine Familie von Methoden, die verschiedenen Zwecken dienen und selten nur messen sollen, sondern meist mit zahlreichen Begleit- und Folgeprozessen zu Verfahren der Organisationsentwicklung erweitert oder sogar Teil der Führungssysteme werden. MABs haben sich auch in den letzten Jahren rasant weiterentwickelt und verändert, weil

aus der Anwendungspraxis ständig neue Anforderungen an sie gestellt wurden, immer neue Chancen gesehen wurden und weil auch die technische Entwicklung neue Wege anbot. MABs sind aber nur dann optimal gestaltbar, wenn der Anwender nicht nur den Befragungsteil gut versteht, sondern auch das gesamte Spektrum der Vorlauf-, Begleit- und Folgeprozesse. Der Workshop deckt daher die MAB-Thematik in ganzer Breite ab.

Die Teilnehmerzahl ist auf 20 Personen begrenzt. Interessenten werden gebeten, sich bis zum 7. Februar 2003 beim ZUMA-Tagungssekretariat anzumelden (workshop@zuma-mannheim.de, Tel. 0621–1246–221). Für die Teilnahme wird ein Beitrag von 100 Euro erhoben. Der Workshop wird von INGWER BORG geleitet.

* * *

Sozio-demographische Variablen in der Umfrageforschung

25. UND 26. MÄRZ 2003

In jedem Survey gibt es einen mehr oder weniger umfangreichen Kanon an sozio-demographischen Variablen. ZUMA ist seit Jahren in eine Standardisierung der Variablen einbezogen. Der Workshop soll die gängigen Variablen hinsichtlich ihrer Aussage und ihrer methodischen Umsetzung vorstellen und diskutieren. Abschließend werden die Demographischen Standards für Deutschland erläutert.

Die Teilnehmerzahl ist auf 20 Personen begrenzt. Interessenten werden gebeten, sich bis zum 14. Februar 2003 beim ZUMA-Tagungssekretariat anzumelden (workshop@zuma-mannheim.de, Tel. 0621–1246–221). Für die Teilnahme wird ein Beitrag von 100 Euro erhoben. Der Workshop wird von JÜRGEN H.P. HOFFMEYER-ZLOTNIK geleitet.

* * *

Einführung in die Clusteranalyse

1. UND 2. APRIL 2003

Der Workshop stellt Konzepte der klassischen Verfahren der Clusteranalyse vor. Der Schwerpunkt liegt dabei auf hierarchischen, agglomerativen und kmeans-Verfahren.

Praktische Anwendungen der Methode werden anhand konkreter Beispiele vorgestellt. Die praktischen Schritte von Clusteranalysen werden am PC mit Hilfe von SPSS und Clustan bzw. ClustanGraphics vorgeführt. Einen weiteren Schwerpunkt bilden Interpretation und Validierung von Ergebnissen. Interessenten sollten mit den üblichen statistischen Verfahren der empirischen Sozialforschung vertraut sein.

Die Teilnehmerzahl ist auf 20 Personen begrenzt. Interessenten werden gebeten, sich bis zum 20. Februar 2003 beim ZUMA-Tagungssekretariat anzumelden (workshop@zuma-mannheim.de, Tel. 0621–1246–221). Für die Teilnahme wird ein Beitrag von 120 Euro erhoben. Der Workshop wird von CORNELIA ZÜLL und MICHAEL WIEDENBECK geleitet.

* * *

Einführung in Datenmanagement und Datenrecodierung

16. UND 17. JUNI 2003

Anhand eines konkreten Datensatzes soll demonstriert werden, wie die Daten vom Fragebogen bis zum analysefähigen Datensatz zu bearbeiten sind über die Schritte: Codeplanerstellung, Vercodung, Datenerfassung, Datensatzerstellung, Recodierung von Daten zu analysefähigen Variablen. Der Schwerpunkt liegt auf der Recodierung von Daten, die am konkreten Datensatz nicht nur demonstriert werden soll, sondern auch geübt werden kann. Die Handhabung geschieht mit den Programmen WORD und SPSS.

Die Teilnehmerzahl ist auf 16 Personen begrenzt. Interessenten werden gebeten, sich bis zum 09. Mai 2003 beim ZUMA-Tagungssekretariat anzumelden (workshop@zuma-mannheim.de, Tel. 0621–1246–221). Für die Teilnahme wird ein Beitrag von 100 Euro erhoben. Der Workshop wird von PETRA HARTMANN (Universität Hamburg) und BEATE BESTMANN (Universität Kiel) geleitet.

Evaluation von Studiengängen - Methodische Aspekte

24. UND 25. JUNI 2003

Der Workshop zur Evaluation von Studiengängen gliedert sich in drei Teile. Zunächst werden Modelle der externen und internen Evaluation von Studienfächern vorgestellt, wie die wechselseitige Evaluation oder die Einrichtung der „auswertenden Konferenz". Im zweiten Teil werden Zielvereinbarungen als follow-up und als Instrument der Qualitätsentwicklung vorgestellt und diskutiert. Der dritte Teil beschäftigt sich mit der Akkreditierung und Zertifizierung von Studiengängen. Hierzu werden Vertreter der AQAS (Agentur für Qualitätssicherung durch Akkreditierung von Studiengängen) und des ACQUIN (Akkreditierungs-, Certifizierungs- und Qualitätssicherungs-Institut) eingeladen.

Die Teilnehmerzahl ist auf 20 Personen begrenzt. Interessenten werden gebeten, sich bis zum 16. Mai 2003 beim ZUMA-Tagungssekretariat anzumelden (workshop@zuma-mannheim.de, Tel. 0621–1246–221). Für die Teilnahme wird ein Beitrag von 100 Euro erhoben. Der Workshop wird von JÜRGEN H.P. HOFFMEYER-ZLOTNIK und HELMUT KROMREY (FU Berlin) geleitet.

* * *

Einführung in die Analyse linearer Strukturgleichungsmodelle mit LISREL 8.52

26. UND 27. JUNI 2003

Das von Karl Jöreskog und Dag Sörbom entwickelte LISREL-Programm ist das bekannteste Statistiksystem zur Analyse von linearen Strukturgleichungsmodellen. In seiner aktuellen Version LISREL 8.52 können konfirmatorische Faktorenanalysen und Pfadmodelle mit beobachteten und latenten Variablen auf der Basis metrischer oder ordinaler Messungen, simultaner Gruppenvergleiche und Analysen von Mittelwertstrukturen berechnet werden, wobei unterschiedliche Schätzmethoden zur Verfügung stehen. Erweitert wurde die Behandlung fehlender Fälle durch die Implementation verschiedener Imputationsverfahren und die ML-Schätzung mit unvollständigen Daten. Neben den klassi-

schen Strukturgleichungsmodellen enthält das Programmsystem auch Module zur explorativen Datenanalyse und zu Mehrebenenmodellen. In dem zweitägigen Workshop wird neben einer Einführung in das Programm auch die Logik der statistischen Modellierung vorgestellt. Es werden keine Vorerfahrungen in der Anwendung linearer Strukturgleichungsmodelle vorausgesetzt.

Die Teilnehmerzahl ist auf 16 Personen begrenzt. Interessenten werden gebeten, sich bis zum 16. Mai 2003 beim ZUMA-Tagungssekretariat anzumelden (workshop@zuma-mannheim.de, Tel. 0621–1246–221). Für die Teilnahme wird ein Beitrag von 150 Euro erhoben. Der Workshop wird von STEFFEN KÜHNEL (Universität Göttingen) geleitet.

* * *

Hermeneutic Procedures to Improve Questionnaire Research

2. UND 3. JULI 2003

This workshop will identify several shortcomings of questionnaires which reduce their information value. Specific hermeneutic procedures will be introduced which can generate more accurate information about individuals' attitudes, personalities, motivation, emotions, and cognitive processes. Participants will get hands-on experience with the procedures by working on exercises during the workshop. The hermeneutic procedures will primarily concern analysis of narrative material such as extended statements by subjects, interview protocols, news accounts, and even literary and historical documents. Some time will also be devoted to a hermeneutic analysis of observed behavior. Criteria for objective interpretation of narratives and behavior will be discussed. The workshop will demonstrate the feasibility of conducting hermeneutic research with large samples of subjects. The workshop will discuss the application of hermeneutic procedures to marketing research, personnel issues, education research, and social science in general.

Participation is limited to 20 people. Prospective participants are asked to apply to the ZUMA conference secretariat by May 23, 2003 (workshop@zuma mannheim.de, Tel. 0621-1246-221). There will be a fee of 120 Euro. The Workshop is led by CARL RATNER (Humboldt State University, California/USA).

‚Political Polls': Politische Umfragen in den Massenmedien

10. UND 11. JULI 2003

Die Bedeutung von Wahlprognosen auf der Grundlage der Ergebnisse von Wahlumfragen erreichte in der politischen Berichterstattung der deutschen Massenmedien bei der vergangenen Bundestagswahl einen neuen Höhepunkt. Angesichts des prognostizierten knappen Wahlausgangs, uneinheitlicher Prognosen, aber auch evidenter Prognosefehler wurde 2002 noch intensiver als bei früheren Wahlen über Qualität, Aussagekraft und Zustandekommen der publizierten Daten diskutiert. In dieser Veranstaltung, die sich primär an Journalisten und andere einschlägig Interessierte richtet, wird am Beispiel der Bundestagswahl 2002 ein Einblick in die Praxis der politischen Umfrageforschung gegeben. Leistungsfähigkeit, aber auch Leistungsgrenzen und Probleme der Meinungsforschung sollen von Referenten aus der Praxis verdeutlicht werden.

Die Teilnehmerzahl ist auf 20 Personen begrenzt. Interessenten werden gebeten, sich bis zum 30. Mai 2003 beim ZUMA-Tagungssekretariat anzumelden (workshop@zuma-mannheim.de, Tel. 0621–1246–221). Für die Teilnahme wird ein Beitrag von 120 Euro erhoben. Der Workshop wird von RÜDIGER SCHMITT-BECK geleitet.

* * *

Survey Design for the Multinational Context

4. UND 5. SEPTEMBER 2003

Starting from the basics of good survey design for a cross-sectional survey in the national context, the two-day workshop goes through the major complicating features to be dealt with in surveys intended for multinational, multilingual implementation. Topics of relevance include sampling design, sample management, contract specifications and contracting, question design and question translation, pre-testing and analysis of item bias, data collection methods, and documentation. Reference will be made to well-known international surveys such as the European Social Survey, the WHO´s World Health Organization Mental Health initiative, the International Social Survey and the Eurobarometer. As the title suggests, the material covered in the workshop focuses on the multinational context and is thus less suited for two- or three-country studies. The workshop is

intended for researchers involved in planning and conducting cross-national non-commercial surveys and will be taught by experts in this field.

Participation is limited to 20 people. Prospective participants are asked to apply to the ZUMA conference secretariat by July 25, 2003 (workshop@zuma mannheim.de, Tel. 0621-1246-221). There will be a fee of 120 Euro. The Workshop is led by JANET HARKNESS.

* * *

Simulation for the Social Sciences

8. BIS 12. SEPTEMBER 2003

This will be a workshop on the techniques of developing simulations to help with the exploration and understanding of social and economic issues. It will provide a rationale for using simulation in the social sciences and outline a number of approaches to social simulation at a level of detail that would enable participants to understand the literature and, for some selected approaches, to develop their own simulations. The workshop covers the basics of modelling and simulation in the social sciences from different points of view (mathematics, computer science, philosophy of science) and of seven different approaches to computer simulation in the social sciences. This workshop is co-organized with ZUMA and will be held at the **University of Koblenz-Landau in: Koblenz, Metternich Campus**. An overview of the module can be found at http://www.uni-koblenz.de/~kgt/SICSS/ModSimNew.html.

The Workshop is led by KLAUS G. TROITZSCH (University of Koblenz-Landau) and NIGEL GILBERT (University of Surrey). Participation is limited to 25 people. Prospective participants are asked to apply to the ZUMA conference secretariat by July 10, 2003 (workshop@zuma mannheim.de, Tel. 0621-1246-221). There will be a fee of 100 Euro.

For all who participated in earlier „Simulation for the social sciences" workshops NIGEL GILBERT and KLAUS G. TROITZSCH will organise an „Advanced Simulation Workshop" to be held at Koblenz from April 8 to 11, 2003. Another such „Advanced Simulation Workshop" is planned for spring 2004. Potential participants will be informed directly from Koblenz.

Analyse linearer Strukturgleichungsmodelle mit LISREL 8.52 für Fortgeschrittene

25. UND 26. SEPTEMBER 2003

Lineare Strukturgleichungsmodelle bieten die Möglichkeit der Analyse komplexer Beziehungsstrukturen. In dem zweitägigen Workshop werden Analysemöglichkeiten vorgestellt, die über die Anwendung einfacher Strukturgleichungsmodelle hinausgehen. Vorgestellt werden Möglichkeiten des simultanen Gruppenvergleichs, die Analyse von Mittelwertstrukturen mit latenten Variablen, wie sie u.a. zur Analyse von Wachstumskurven verwendet werden, die Analyse von Modellen auf der Basis ordinaler Indikatoren und die Berücksichtigung von Interaktionseffekten bei latenten Prädiktoren. Die Umsetzung der Modell-Logik erfolgt mit dem Programm LISREL 8.52. Vorausgesetzt werden Grundkenntnisse der Methodik linearer Strukturgleichungsmodelle.

Die Teilnehmerzahl ist auf 16 Personen begrenzt. Interessenten werden gebeten, sich bis zum 1. August 2003 beim ZUMA-Tagungssekretariat anzumelden (workshop@zuma-mannheim.de, Tel. 0621–1246–221). Für die Teilnahme wird ein Beitrag von 150 Euro erhoben. Der Workshop wird von STEFFEN KÜHNEL (Universität Göttingen) geleitet.

* * *

ADM-Stichprobensysteme in der Umfragepraxis: CATI und face-to-face

30. SEPTEMBER UND 01. OKTOBER 2003

Es werden die ADM-Auswahlgrundlagen für CATI und face-to-face (neues System ab 2003) vorgestellt. Der Einsatz in der Praxis geschieht über die Schritte: Erstellung eines Allokationstableaus (Schichtungsmerkmale bei unterschiedlich großen Stichproben); systematische Zufallsauswahl innerhalb der Zellen; Auswahl der Zielperson im Haushalt (Schwedenschlüssel oder Last-Birthday). Der Einsatz von face-to-face mit Begehung stellt eine mehrstufige Zufallsauswahl für Haushalte und eine Auswahl der Zielperson im Haushalt dar. Abschließend folgen: Ausschöpfungskontrolle und Feldbericht.

Die Teilnehmerzahl ist auf 20 Personen begrenzt. Interessenten werden gebeten, sich bis zum 1. August 2003 beim ZUMA-Tagungssekretariat anzumelden (workshop@zuma-mannheim.de, Tel. 0621–1246–221). Für die Teilnahme wird ein Beitrag von 100 Euro erhoben. Der Workshop wird von JÜRGEN H.P. HOFFMEYER-ZLOTNIK und KURT BEHRENS (BIK Hamburg) geleitet.

3. NUTZERKONFERENZ „FORSCHUNG MIT DEM MIKROZENSUS: ANALYSEN ZUR SOZIALSTRUKTUR UND ZUM ARBEITSMARKT“

9. - 10. OKTOBER 2003, ZUMA, MANNHEIM

CALL FOR PAPERS

In den empirischen Sozial- und Wirtschaftswissenschaften können Scientific-Use-Files des Mikrozensus seit dem Erhebungszeitpunkt 1989 genutzt werden. Bis Ende 2002 werden voraussichtlich die Daten von 1998 und 2000 zur Verfügung stehen (nähere Informationen hierzu: http://www.gesis.org/Dauerbeobachtung/Mikrodaten/Aktuelles/).

Die Konferenz wendet sich an Forscher, die mit den Scientific-Use-Files des Mikrozensus arbeiten oder daran interessiert sind. Ziele der dritten Nutzerkonferenz sind die auf der Basis dieser Daten gewonnenen neueren Forschungsergebnisse zu diskutieren, wie auch den Erfahrungsaustausch der Nutzer untereinander und mit den statistischen Ämtern als Datenproduzenten zu fördern.

Beiträge können aus allen Arbeitsgebieten eingereicht werden, die sich auf der Basis von Scientific-Use-Files des Mikrozensus inhaltlich oder methodisch mit Analysen zur Sozialstruktur und zum Arbeitsmarkt beschäftigen. Folgende Themenschwerpunkte sind vorgesehen:

- Entwicklung und Probleme des Arbeitsmarktes in Ost- und Westdeutschland
- Soziale Lagen und wirtschaftliche Situationen von Familien und Haushalten
- Europäische und internationale Vergleiche
- Nutzung der Ergänzungs- und Zusatzprogramme (z.B. Migranten- und Pendlermerkmale, Pflegeversicherung, Retrospektivfragen)
- Methodische Probleme der Sekundäranalyse des Mikrozensus

Interessenten werden gebeten, Abstracts der geplanten Beiträge bis spätestens 28. Februar 2003 einzusenden. Sie sollten nicht mehr als 1-2 Seiten umfassen sowie den Titel des Beitrags und Angaben über den Verfasser enthalten. Die Abstracts können per Formular

im WWW (siehe http://www.gesis.org/Dauerbeobachtung/Mikrodaten/Veranstaltungen/3Nutzerkonferenz_Einsendungabstracts.htm), per email oder auf Diskette eingereicht werden.

Durchgeführt und organisiert wird die Konferenz von der Abteilung Mikrodaten, ZUMA, und vom Statistischen Bundesamt, Gruppe IX B – Mikrozensus. Ihre Ansprechpartnerin bei ZUMA ist:

Dr. Annette Kohlmann
ZUMA
Postfach 12 21 55
D-68072 Mannheim
Tel.: 0621-1246-253/-265; Fax: 0621-1246-100
http://www.gesis.org/Dauerbeobachtung/Mikrodaten/
kohlmann@zuma-mannheim.de