

GESIS

ZUMA

52

Mai 2003

NACHRICHTEN

**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 27. 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, 2003



In eigener Sache.....	5
-----------------------	---

FORSCHUNGSBERICHTE

Prognosen als Methode zur Reduktion der Offensichtlichkeit von Umfragebefunden: Ein Pilotexperiment. <i>Ingwer Borg & Carola Hillenbrand</i>	7
Computerunterstütztes Pretesting von CATI-Fragebögen: Das CAPTIQ-Verfahren <i>Frank Faulbaum, Marc Deutschmann & Martin Kleudgen</i>	20
Animaqu – eine computerbasierte Befragung als Möglichkeit zur Erfassung besonderer Zielgruppen. Ein Beispiel einer Anwendung bei gehörlosen Menschen. <i>Joachim Gerich, Roland Lehner, Johannes Fellingner & Daniel Holzinger</i>	35
Umfrageeinstellung und Umfrageerfahrung. Die relative Bedeutung unterschiedlicher Aspekte der Interviewererfahrung für die generalisierte Umfrageeinstellung. <i>Volker Stocké & Bettina Langfeldt</i>	55
Bildungspartizipation und Heiratsneigung: Die Entwicklung des bildungsselektiven Heiratsverhaltens in Westdeutschland zwischen 1970 und 1997. <i>Heike Wirth & Simone Schmidt</i>	89

MITTEILUNGEN

Berufsklassifikation und Messung des beruflichen Status/Prestige. <i>Jürgen H.P. Hoffmeyer-Zlotnik & Alfons Geis</i>	125
Study Monitoring in the ISSP. <i>Evi Scholz & Sabine Klein</i>	139
Mikrodaten (German Microdata Lab): Das Servicezentrum für amtliche Mikrodaten bei ZUMA. <i>Paul Lüttinger, Bernhard Schimml-Neimanns, Heike Wirth & Georg Papastefanou</i>	153

ZUMA-PUBLIKATIONEN

ZUMA-Methodenberichte	173
Buchhinweise.....	174

ZUMA-VERANSTALTUNGEN 2003

Übersicht über die Veranstaltungen 2003	178
Programm der 3. Nutzerkonferenz „Forschung mit dem Mikrozensus“, 9.-10. Oktober	179

DURCHWAHL-RUFNUMMERN

IN EIGENER SACHE

Den Sozialwissenschaften wird häufig unterstellt, dass die Ergebnisse ihrer Forschung offensichtlich sind. Das ist insbesondere für die Umfrageforschung ein bislang wenig beachtetes Problem, weil ihre Ergebnisse unter dem Eindruck des „Habe ich sowieso schon gewusst“, oder „Das hätte ich Ihnen alles vorher schon sagen können“ kaum noch die nötige kritische Aufmerksamkeit bekommen. INGWER BORG & CAROLA HILLENBRAND untersuchen eine Methode, die den Eindruck der Offensichtlichkeit von Umfragebefunden reduzieren soll.

Der Beitrag von *FRANK FAULBAUM, MARC DEUTSCHMANN & MARTIN KLEUDGEN* stellt ein computerunterstütztes, auf dem Prinzip des Behavior-Coding basierendes Pretestverfahren für CATI-Instrumente unter Feldbedingungen vor. Diese sogenannte CAPTIQ-Verfahren (Computer Assisted Pretesting of Telephone Interview Questionnaires) erlaubt die sofortige computerunterstützte Codierung der Problemarten und liefert einen Datensatz, auf dessen Grundlage die Pretestergebnisse mit Hilfe einer grafischen Darstellung sichtbar gemacht werden können.

Die Anwendung einer schriftlichen standardisierten Befragung ist an bestimmte Voraussetzungen der interviewten Personen gebunden. Insbesondere ist eine ausreichende Lesekompetenz eine zentrale Anwendungsvoraussetzung. Für gehörlose Menschen ist die Gebärdensprache die eindeutig bevorzugte Kommunikationsform. Die Lesekompetenz liegt im Durchschnitt unter der der Normalbevölkerung und weist eine große Variationsbreite auf. *JOACHIM GERICH, ROLAND LEHNER, JOHANNES FELLINGER & DANIEL HOLZINGER* stellen eine computergestützte Erhebungsförm vor, welche die selbstadministrierte Befragung dieser Zielgruppe ermöglicht.

VOLKER STOCKÉ & BETTINA LANGFELDT untersuchen die Frage, ob sich die Bewertung vergangener Interviewverfahren auf die Einstellungen gegenüber Umfragen auswirkt. Dabei steht insbesondere die Frage nach der relativen Erklärungskraft verschiedener Bewertungsdimensionen dieser Erfahrung im Mittelpunkt.

Die Höherqualifizierung von Frauen und ihre hierdurch verbesserten Berufschancen zählen zu den zentralen Faktoren des sozialen Wandels. In der aktuellen gesellschafts- und sozialwissenschaftlichen Diskussion wird dies als ein Grund für den Wandel in Familiengründungsprozessen, wie etwa dem Rückgang der Heirats- und Geburtenhäufigkeit gesehen. Der Beitrag von *HEIKE WIRTH & SIMONE SCHMIDT* greift die Fragen auf, inwieweit die Höherqualifizierung von Frauen zu einer Veränderung der

Heiratsneigung geführt hat und ob ein vergleichbarer Zusammenhang zwischen dem individuellen Bildungsniveau von Männern und dem Haushaltsmodus (Single, nichteheliche Lebensgemeinschaft versus Verheiratete) besteht. Für die Analysen werden die Volkszählungsdaten von 1970 und des Mikrozensus von 1997 genutzt.

Bei den Mitteilungen möchte ich insbesondere auf den Beitrag von PAUL LÜTTINGER, BERNHARD SCHIMPL-NEIMANN, HEIKE WIRTH & GEORG PAPASTEFANOU hinweisen. Er beschreibt die Aufgaben und Ziele des German Microdata Lab bei ZUMA, das am 1. Juli offiziell seine Arbeit aufnimmt.

Neu bei ZUMA sind NINA ROTHER (PIONEUR-Projekt: Pioneurs of Europe's Integration from below) und LARS KACZMIREK (WebSM-Projekt: Web Survey Methodology Site). Ihre Tätigkeit bei ZUMA beendet haben MICHAEL BOSNJAK, CHRISTINA WEBER und SIMONE SCHMIDT. RÜDIGER SCHMITT-BECK hat den Ruf auf eine C3 Professur an die Universität Duisburg angenommen, wozu wir ihm herzlich gratulieren.

Zum Schluss noch zwei redaktionelle Hinweise: In der letzten Ausgabe wurde versehentlich PETER PH. MOHLER als Herausgeber der ZUMA-Nachrichten benannt. Die Herausgeberschaft liegt aber beim Vorsitzenden des ZUMA e.V. Des weiteren verzeichnen wir einen Wechsel in der Redaktion. PAUL LÜTTINGER beendet nach 30 Ausgaben seine redaktionelle Tätigkeit für die ZUMA-Nachrichten. Ihm gebührt ein besonderer Dank für seine sicherlich nicht immer einfache Aufgabe.

Jan van Deth
Vorsitzender des ZUMA e.V.

PROGNOSEN ALS METHODE ZUR REDUKTION DER OFFENSICHTLICHKEIT VON UMFRAGEBEFUNDEN: EIN PILOTEXPERIMENT

INGWER BORG & CAROLA HILLENBRAND¹

Untersucht wird hier eine Methode, die den Eindruck der Offensichtlichkeit von Umfragebefunden reduzieren soll. Versuchspersonen (Vpn) wurden gebeten, die Ergebnisse für verschiedene Items aus angenommenen Umfragen vorherzusagen bzw. sie bekamen die Vorhersagen anderer Personen gezeigt. Es zeigte sich, dass diese Maßnahmen dazu führten, dass die Vpn Umfragebefunde als überraschender und als interessanter bewerten als Personen der Kontrollgruppe. Keinen systematischen Einfluss haben die Prognosen auf die Bewertung der praktischen Nützlichkeit der Befunde.

This paper investigates whether prognoses on the results of surveys help to reduce the common impression of obviousness when encountering survey findings. An experiment was conducted where subjects were asked to predict the results of a number of fictitious survey items and/or they were shown predictions made by other people on these items. It was found that making and/or seeing such prognoses indeed makes survey findings more unexpected and more interesting. On the other hand, prognoses have no systematic effect on the subjects' ratings of the practical usefulness of the findings.

1. Einleitung

Den Sozialwissenschaften wird häufig unterstellt, dass das, was sie mit ihrer Forschung herausfinden, offensichtlich ist. Das ist insbesondere für die Umfrageforschung ein Problem, weil ihre Ergebnisse unter dem Eindruck des „Habe ich sowieso schon gewusst“, des

¹ Carola Hillenbrand ist Research Fellow am Henley Management College Greenlands, Henley-on-Thames.

„Das hätte ich Ihnen alles vorher schon sagen können“ kaum noch die nötige kritische Aufmerksamkeit bekommen.

Die scheinbare Offensichtlichkeit sozialwissenschaftlicher Befunde hängt eng zusammen mit dem Hindsight Bias [auch: I-knew-it-all-along-Effekt (Stahlberg u.a. 1993) oder Rückschaufehler (Pohl 1992)]. Hierzu gibt es zahlreiche Studien. So berichtet Baratz (1983) z.B. ein Experiment mit Studenten, denen sie vermeintliche Forschungsbefunde (z.B. „Personen, die regelmäßig zur Kirche gehen haben statistisch mehr Kinder als Personen, die selten zur Kirche gehen“) bzw. deren Gegenteil („Personen, die ... haben statistisch weniger ...“) präsentierte. Die Vpn gaben jeweils an, dass der Befund sie nicht überrascht, sondern dem entspricht, was sie vorausgesagt hätten. Das Gefühl, schon alles gewusst zu haben, geht einher mit der Rückschautäuschung. Es tritt interessanterweise verstärkt auf bei Experten: „Oft zeigt sich Expertentum mehr in übersteigter Selbstsicherheit („overconfidence“) als in einer objektiv ansteigenden Genauigkeit und Leistung bei Entscheidungen und Prognosen“ (Fiedler 1993: 7).

Die Rückschautäuschung und die damit einhergehende Overconfidence haben zu zahlreichen Erklärungen geführt. Die älteste Theorie ist die von Fischhoff (1983), der das Phänomen als Folge einer sofortigen Assimilation neuer Informationen in vorhandene Wissensstrukturen, als Echtzeit-Updating der Wissensbasis durch neue Informationen interpretiert. Demnach würde die Rückschau auf diesen bereits veränderten Strukturen basieren. Stahlberg u.a. (1995) argumentieren dagegen, dass Personen, die sich nicht an ihr ursprüngliches Urteil erinnern können (und nur bei diesen soll der Hindsight Bias auftreten), die nachträgliche Information als Anker für die Rekonstruktion des ursprünglichen Urteils verwenden. Campell/Tesser (1983) verweisen zudem auf motivationale Faktoren (Selbstwertgefühl, soziale Erwünschtheit, Bedürfnis nach Kontrolle), denen zumindest die öffentliche Darstellung des Urteils unterliegen könnte.

Das „Problem der Offensichtlichkeit“ (Lazarsfeld 1949) ist in der Literatur wiederholt bemerkt worden. In einem genialen Review des monumentalen „The American Soldier“ hat Lazarsfeld (1949) es wie folgt genutzt. Er berichtet zunächst einige Befunde dieser Studie, so z.B.:

- Soldaten mit besserer Bildung hatten größere Anpassungsprobleme im Krieg als Soldaten mit geringer Bildung.
- Soldaten aus den Südstaaten kamen mit dem heißen Südseeklima besser zurecht als Nordstaatler.
- So lange der Krieg andauerte waren die Soldaten motivierter nach Hause zurückzukehren als nach dem Krieg.

Alle Befunde erscheinen offensichtlich und leicht erklärbar. Nachdem der Leser diese abgenickt hat, klärt ihn Lazarsfeld über die *wirklichen* Befunde auf: Soldaten mit besserer Bildung hatten geringere Anpassungsprobleme als weniger Gebildete; Südstaatler kamen mit dem Südpazifiklima nicht besser zurecht als Nordstaatler; und die Rückkehrmotivation war nach dem Krieg größer als während des Krieges. Lazarsfeld merkt dazu an, dass wenn er diese Befunde berichtet hätte, der Leser dann eben diese als offensichtlich bezeichnet hätte.

Den Trick, den Rezipienten des Forschungsergebnisses auf eine Aussage festzulegen, bevor er das tatsächliche Ergebnis sieht, kann man als Versuch zum Debiasing, zur Enttäuschung bezeichnen. Lazarsfeld hat damit ein Forschungsgebiet antizipiert, das fragt, wie man kognitive Täuschungen verringern oder eliminieren kann. Fischhoff (1977, 1983) hat sich hiermit als erster beschäftigt, wenn auch nur mit der Absicht, auf diese Weise die Robustheit des Hindsight Bias zu testen. Von ihm und anderen wurden verschiedene Debiasing-Methoden vorgeschlagen und getestet, wie z.B. das Zurücknehmen von berichteten Ergebnissen als „falsch“ (Hasher u.a. 1981); die Präsentation von „sehr überraschenden Feedback-Informationen“ (Kahneman 1992); Manipulationen der Glaubwürdigkeit der Feedbackinformation (Pohl 1998); Manipulationen der Kosten suboptimaler Urteile (Arkes 1991) bzw. der Incentives für optimale Urteile (Hell u.a., 1988); oder eine vollständige Aufklärung über den Hindsight Bias (Fischhoff 1977; Pohl/Hell 1996).

Alle diese Techniken dienen dazu, die Genauigkeit der rekonstruierten Urteile zu verbessern. Für die Reduktion der scheinbaren Offensichtlichkeit von Umfragebefunden in der Umfragepraxis erscheinen sie wenig geeignet. Folgt man Fischhoffs (1983) Theorie, dann müsste man verhindern, dass diese Befunde sofort und unbemerkt in die Wissensbasis der Zielpersonen eingehen. Dafür ließe sich evtl. die „Consider the Opposite“-Technik anwenden, in der die Zielperson einen Hinweis zu einer eher „bedächtigen“ Rezeption der Befunden bekommt, der wie folgt lauten könnte: „Ask yourself at each step whether you would have made the same evaluation had exactly the same study produced results on the other side of the issue“ (Lord u.a., 1984: 1233). Ebenso könnte man sie auch auffordern, für jeden Befund möglichst viele und gute Argumente zu suchen, die apriorisch ein anderes Ergebnis nahe legen („counterfactual reasoning“). Diese Strategien reduzieren i. Allg. erfolgreich den Hindsight Bias.

Die Technik von Lazarsfeld erscheint von allen Ansätzen am besten geeignet für den praktischen Einsatz. Sie lässt sich so verwenden, dass man für einen Teil der Items schriftlich Prognosen erhebt und diese dann später mit den tatsächlichen Ergebnissen vergleicht. Borg (2002, 2003) hat dieses Verfahren im Rahmen von Mitarbeiterbefragungen verwendet. Dort wurden Führungskräfte gebeten, vorherzusagen,

zu welchen Prozentsatz ihre Mitarbeiter einigen ausgewählten Items jeweils zustimmen. Die mittleren Prognosen wurden später mit den empirischen Zustimmungsteilen verglichen. Eine Variante dabei war, auch den Range der Prognosen oder sogar ihre gesamte Verteilung zu zeigen. Dieses Feedback führte dazu, dass erhebliche Diskussionen entstanden, weil einige der Prognosen immer ganz falsch waren und weil der Range der Prognosen immer beträchtlich war. Die Führungskräfte lernten so schnell, dass sie zumindest uneins waren in der Wahrnehmung der Einstellungen und Meinungen der Mitarbeiter. Zudem zeigten sich einige inhaltliche Auffälligkeiten. So sagten die Führungskräfte z.B. die Zufriedenheit der Mitarbeiter bei den Themen, mit denen sie selbst relativ zufrieden waren (z.B. Bezahlung), viel zu hoch voraus. Derartige Befunde führten zu zusätzlichen Diskussionen und zu einer scheinbar offeneren und interessierteren Einstellung den Umfrageergebnissen gegenüber.

Diese Beobachtungen aus der Praxis sollen im folgenden experimentell überprüft werden. Wir fragen also, ob die Abgabe von Prognosen und verschiedene Formen der Rückmeldung von Prognosen tatsächlich dazu führen, dass Vpn Umfragebefunde als interessanter und informativer beurteilen. Wir fragen weiter, ob auch die Prognosen von anderen Personen zu ähnlichen Effekten führen oder ob es erforderlich ist, dass man selbst Prognosen abgibt. Zudem wollen wir wissen, ob verschiedene Formen der Rückmeldung – mittlere Prognosewerte, Range der Prognosen – zu unterschiedlichen Effekten führen.

2. Methode

2.1 Versuchspersonen

Als Versuchspersonen (Vpn) dienten 125 Personen aus dem Raum Augsburg. Unter den Vpn waren keine Studenten. Das Alter der Vpn lag zwischen 17 und 65 Jahre, mit einem Mittel von 35. Die Verteilung der Geschlechter war 65 Frauen und 60 Männer. Die Stichprobe war eine Gelegenheitsstichprobe, die über Aushänge und Handzettel rekrutiert wurde. Als Incentive wurden den Vpn die Teilnahme an einem bayrischen Weißwurstfrühstück bzw. an Kaffee und Kuchen angeboten. Die Vpn wurden per Zufall einer von fünf Gruppen (siehe unten) zugewiesen.

2.2 Unabhängige Variablen

Vor den Treatments wurden zwei Fragen gestellt: „Für wie sinnvoll halten Sie psychologische Forschungsarbeit?“ (mit einer Ratingskala von 1=„nicht sinnvoll“ bis 6=„sehr sinnvoll“) und „Halten Sie es für schwierig, die Ergebnisse sozialwissenschaftlicher Umfragen vorherzusagen?“ (mit einer Ratingskala von 1=„sehr einfach“ bis 6=„sehr schwierig“).

Das Treatment wurde eingeleitet mit der Frage: „Auf den folgen Seiten werden Ihnen zwölf interessante Fragen aus aktuellen Umfragen vorgestellt.“ (Diese Fragen waren in Pretests überprüft worden in Bezug auf ihre Interessanztheit und Realitätsnähe.) Die Fragen lauteten z.B.: „In einer repräsentativen Umfrage wurden ADAC-Mitglieder befragt, ob sie die Einführung einer Autobahngebühr befürworten“ und „In einer repräsentativen Umfrage wurden deutsche Steuerzahler gefragt, ob sie bei der Einkommenssteuererklärung schummeln.“

Die Vpn hatten dann, je nach Versuchsgruppe, zunächst folgende Aufgabe. (a) Ohne Prognosen: „Zu Beginn sehen Sie die Fragestellung der jeweiligen Untersuchung. Daran schließt sich das Ergebnis und die graphische Darstellung des Ergebnisses an. Bitte lesen Sie sich die Ergebnisse aufmerksam durch!“ Oder (b) mit Prognosen: „Wir wollen von Ihnen wissen, welches statistische Ergebnis Sie für jede dieser Fragen erwarten. Hierbei sollen Sie angeben, wie viel Prozent der in der repräsentativen Umfrage befragten Personen den jeweiligen Aussagen zugestimmt haben.“ Diese Urteile wurden für Bedingung (b) sequentiell Item für Item erhoben.

Hieran anschließend bekamen die Vpn entweder keine Rückmeldung über evtl. Prognosewerte oder eines der beiden Treatments. (1) „Vor Ihnen hat bereits eine ähnliche Gruppe von Personen Vorhersagen abgegeben, zu welchen Ergebnissen die Umfragen ihrer Meinung nach führen. ... Im Folgenden zeigen wir Ihnen die durchschnittlich vorhergesagten Werte dieser Vergleichsgruppe. Im Anschluss daran stehen die tatsächlichen Umfrageergebnisse und eine graphische Darstellung der Ergebnisse“. Oder (2): „Vor Ihnen hat bereits... Dabei kam es teilweise zu recht unterschiedlichen Vorhersagen, Im Folgenden zeigen wir Ihnen zu jeder Fragestellung den kleinsten und den größten Wert, der von den Personen der Vergleichsgruppe vorhergesagt wurde...“.

Das Design weist also fünf Experimentalgruppen mit jeweils 25 Vpn auf:

Gruppe	Prognosen	Rückmeldung von Prognosen
1	keine	keine
2	eigene	keine
3	keine	mittlere Prognosen
4	eigene	mittlere Prognosen
5	eigene	Range der Prognosen (Min-Max)

2.3 Abhängige Variablen

Die Vpn sollten danach die Umfragebefunde bewerten auf den folgenden sechs-stufigen, bipolaren Skalen: „spannend vs. langweilig“, „nutzlos vs. hilfreich“, „offensichtlich (war ja klar) vs. latent (hätte ich nicht vermutet)“, „unsinnig vs. sinnvoll“, „altbekannt vs. neuartig“, „interessant vs. uninteressant“, „informativ vs. irrelevant“, „abgehoben vs. bodenständig“, „unauffällig vs. beachtenswert“, „lebensnah/praktisch vs. theoretisch“ und „überraschend vs. erwartet“.

Danach wurden die Vpn gebeten, folgende Fragen zu beantworten: „Was schätzen Sie, in wie viel Prozent der vorgestellten Umfrageergebnisse hätten/haben Sie in etwa eine richtige Prognose abgegeben?“; „Für wie sinnvoll halten Sie psychologische Forschungsarbeit?“; „Halten Sie es für schwierig, die Ergebnisse sozialwissenschaftlicher Umfragen vorherzusagen?“ (Die beiden letzten Fragen wiederholen die Fragen von oben.)

In einer Unterbrechungsphase sollten die Vpn die Grandiositätsskala ausfüllen (siehe unten). Schließlich wurde ihnen noch folgende Aufgabe unterbreitet: „Abschließend notieren Sie bitte kurz, welche von den dargestellten Umfragethemen Ihnen noch in Erinnerung sind. Geben Sie dabei auch die tatsächlichen (nicht die von Ihnen vorhergesagten) Umfrageergebnisse an, soweit Sie sich daran erinnern können.“

2.4 Moderatorvariablen

Eine naheliegende Vermutung ist die, dass der Eindruck der Offensichtlichkeit auch von bestimmten persönlichkeitspsychologischen Dispositionen abhängt, insbesondere von Narzissmus und Grandiosität. Um dies zu prüfen, wurde die von Meyer (1999, 2000) beschriebene Grandiositätsskala verwendet, ein kurzes 11-Item-Instrument. Sie enthält Items wie „Ich würde mich als außergewöhnlichen Menschen beschreiben“, „Ich habe eine Menge Qualitäten, die mich von anderen Menschen abheben“ und „Ich spüre häufig die Bewunderung anderer Menschen“. Dazu schreibt Meyer (1999: 8): „Die Items entsprechen damit konzeptuell der Facette ‚Grandiosität‘ der narzisstischen Persönlichkeitsstörung im DSM-IV. Die Skala kovariiert substantiell positiv mit Bewunderung, narzisstisch sozialen Kognitionen und Exhibit-Eitelkeit und ist unabhängig von der Skala Sensitivität. Ein starker positiver Zusammenhang besteht zur Extraversion ($r=.42$) und der Rosenberg Selbstwertskala ($r=.37$). Die Skala ist negativ korreliert mit Impression-Management“.

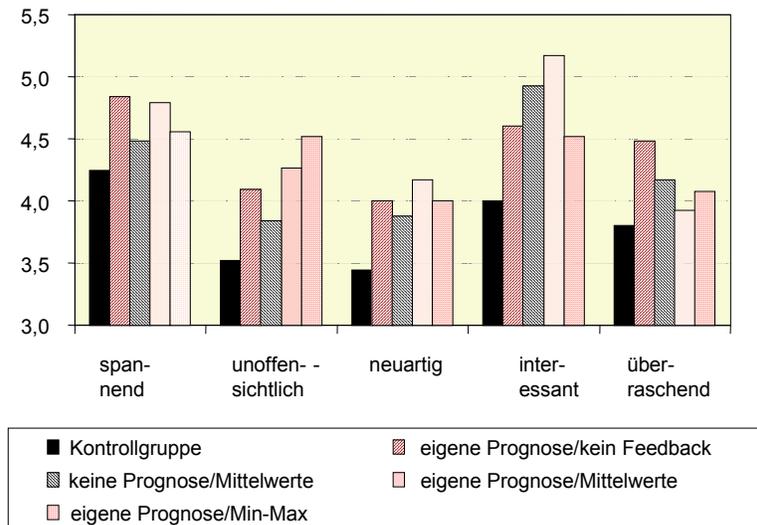
3. Ergebnisse

Eine MDS-Analyse (Borg/Groenen 1997) zeigt, dass die 125 Vpn die Beurteilungskriterien in zwei Gruppen strukturieren. Diese umfassen zum einen die Kriterien Latent (i.S. von „hätte ich nicht vermutet“), Interessant, Überraschend, Spannend und Neuartig zum

anderen die Skalen Sinnvoll, Praktisch, Hilfreich, Informativ-Relevant und Bodenständig. Das Attribut Beachtenswert hat keine klare Zuordnung und liegt in der MDS-Ebene zwischen den Clustern. Die beiden Eigenschaftscluster kann man als Unerwartet-Interessant bzw. als Nützlich bezeichnen.

Die Bewertung der präsentierten Umfragebefunde bei den Kriterien des Clusters Unerwartet-Interessant zeigt Abbildung 1. Wie man sieht, beurteilen alle Gruppen, die irgendein Treatment bekamen (eigene Prognosen, Rückmeldung der Prognosen anderer oder beides) die Umfragebefunde in jedem der 20 Fälle im Mittel positiver als die Vpn der Kontrollgruppe ($p=0,000$ nach dem Vorzeichentest). Insbesondere erkennt man, dass die Umfragebefunde nach den Treatments als vergleichsweise weniger offensichtlich beurteilt werden. Die Art der Treatments spielt offenbar keine systematische Rolle. Insofern zeigt Abbildung 1, dass die Treatments sich dahingehend ausgewirkt haben, dass die Umfrageergebnisse als unerwarteter, interessanter und überraschender erlebt werden.

Abbildung 1: Bewertung von Umfragebefunden auf den Kriterien des Clusters Unerwartet-Interessant



Bei den Kriterien des Clusters Nützlich zeigt sich dagegen kein systematischer Effekt der Treatments (Abbildung 2). Hier sind die Skalenmittelwerte der Treatmentgruppen in neun Fällen größer, in zehn Fällen kleiner und einem Fall gleich groß wie der entsprechende Skalenmittelwert der Kontrollgruppe. Bei keiner der Skalen sind alle Mittelwerte der Treatmentgruppen größer als der Mittelwert der Kontrollgruppe. Zudem ist der Range der Mittelwerte kleiner als beim Cluster Unerwartet-Interessant.

Abbildung 2: Bewertung der Umfragebefunde auf den Kriterien des Clusters „nützlich“

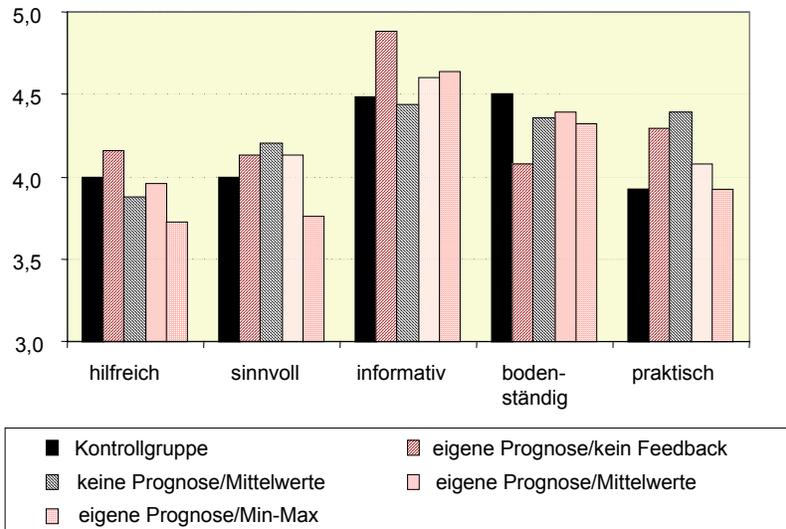
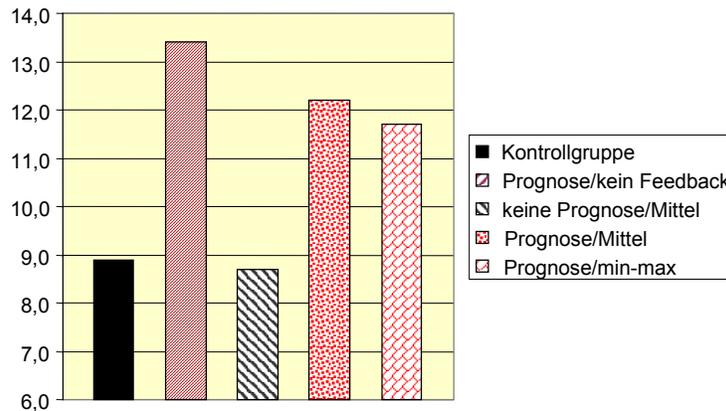


Abbildung 3 zeigt, dass das Treatment „eigene Prognosen“ auf eine andere abhängige Variable einen einfachen Effekt hat: Wie man sieht werden die Umfragebefunde besser erinnert, wenn die Vpn eigene Prognosen abgegeben haben. Der Erinnerungsskore wurde dabei so konstruiert, dass die Vpn einen Punkt dafür bekamen, wenn sie den Inhalt des Items erinnern konnten und einen zusätzlichen Punkt, wenn sie den Umfragewert innerhalb von plus/minus 5 Prozentpunkten richtig wiedergeben konnten. Wie man sieht, sind die Erinnerungswerte deutlich (und signifikant i.S. von t-Tests) besser für die Gruppen, die eigene Prognosen abgegeben haben.

Abbildung 3: Mittlere Erinnerungswerte für die Umfrageitems unter fünf Bedingungen



Systematische Effekte der Treatments zeigen sich auch beim Vergleich der Antworten auf die allgemeinen Fragen „Für wie sinnvoll halten Sie psychologische Forschungsarbeit?“ und „Halten Sie es für schwierig, die Ergebnisse sozialwissenschaftlicher Umfragen vorherzusagen?“, die ganz am Anfang und nochmals ganz zum Schluss des Experiments gestellt wurden. Während bei der Kontrollgruppe die mittleren Ratings nur um 0,06 bzw. um 0,03 Skaleneinheiten ansteigen, sind die Veränderungen bei den vier Treatmentgruppen deutlich größer: Beim Kriterium „sinnvoll“ wachsen bei den Treatmentgruppen die Ratings um 0,44*, 0,48*, 0,64** bzw. 0,64* Einheiten; beim Kriterium „schwierig“ wachsen sie um 0,68*, 0,20, 0,40** bzw. 0,32 Einheiten (paarweise t-Tests). Die Vpn in den Treatmentgruppen meinen also am Ende des Experiments verstärkt, dass es schwierig ist, Umfrageergebnisse vorherzusagen. Ebenso führen die Treatments dazu, dass psychologische Forschung ganz allgemein als sinnvoller bewertet wird.

Keine systematische Beziehung zu den Bewertungen oder zu den Erinnerungswerten haben dagegen die Grandiositätsscores der Vpn. Sie korrelieren lediglich mit dem Rating, wie schwer es ist, Umfragedaten i. Allg. vorherzusagen, signifikant. Die Korrelationen sind $r=-0,22^*$ vor den Treatments und $r=-0,27^*$ nach den Treatments. Vpn mit hohen Grandiositätswerten sind also allgemein der Meinung, dass es relativ leicht ist, Umfragedaten vorherzusagen. Diese Überzeugung steigt nach dem Experiment noch etwas an, obwohl sie die tatsächlichen Umfragebefunde auch nicht anders – insbesondere auch nicht als weniger unoffensichtlich – bewerten als Personen mit geringen

Grandiositätswerten. Allerdings erweist sich die Grandiositätsskala für die 125 Vpn als wenig homogen ($\text{Alpha}=0,71$). Subskalen bzw. die Einzelitems korrelieren allerdings alle ähnlich wie die Gesamtskala mit den anderen Variablen. Sie korrelieren insbesondere alle negativ mit der Einschätzung, wie schwierig die Vorhersage von Umfragedaten ist.

4. Diskussion

Wir haben hier gezeigt, dass es möglich ist, die Offensichtlichkeit von Umfragebefunden durch die Verwendung von Prognosen zu reduzieren. Genauer gesagt geben Personen, die eigene Prognosen abgegeben haben, aber auch solche, die nur die Prognosen anderer gesehen haben, in stärkerem Ausmaß an, dass sie die Umfragebefunde überraschend, interessant und nicht offensichtlich finden. Die praktische Relevanz der Umfragebefunde wird dagegen von den verschiedenen Gruppen nicht systematisch verschieden bewertet. Prognosen führen zudem dazu, dass Umfragebefunde ganz allgemein als schwieriger vorhersagbar und psychologische Forschung allgemein als sinnvoller eingeschätzt wird.

Ein nicht erwarteter Befund ist, dass die Abgabe eigener Prognosen zwar die Erinnerung an die Items und die Umfragebefunde verbessert, aber nicht dazu führt, dass die Umfragebefunde als interessanter bewertet werden als in dem Fall, in dem die Vpn keine eigenen Prognosen abgeben, sondern nur die Prognosen anderer sehen. Überlegenswert erscheint aber, ob hier nicht doch ein Effekt dann entstehen könnte, wenn man den Personen ihre jeweils eigenen, individuellen Prognosen und nicht irgendwelche Mittelwerte zurückspegelt. Allerdings erscheint es möglich, dass dieser Ansatz unter gewissen Bedingungen auch kontraproduktiv sein kann. So zeigen Sanna u.a. (2002), dass auch die Methode des „consider the opposite“ dann nicht zu einem Debiasingeffekt führt, wenn die Personen es schwer finden, Argumente dafür zu finden, dass die berichteten Befunde wenig wahrscheinlich sind. Die Prognosetechnik basiert in gewissem Umfang darauf, dass einige der Prognosen falsch sind oder dass, im Fall von Fremdprognosen, zumindest der Range der Prognosen groß ist. Ist das nicht so, besteht kein theoretischer Grund, warum Prognosen die Offensichtlichkeit der Befunde reduzieren sollten. Nur im Fall eigener Prognosen könnte eine Offensichtlichkeitsreduktion auch dadurch entstehen, dass den Personen das Abgeben der Prognosen erlebbar schwer fällt, dass sie also nicht so leicht „verfügbar“ sind.

Das Persönlichkeitsmerkmal Grandiosität führt zu keinen besonderen Einsichten. Das kann allerdings daran liegen, dass diese Skala mit Studenten als Probanden entwickelt wurde. Die Vpn in dieser Studie waren aber keine Studenten. Für weitere Experimente wäre es daher von Interesse, eher bewährte Instrumente zur Messung der Persönlichkeit zu verwenden, weil unserer Meinung nach die Hypothese, dass die Bewertung von Um-

fragebefunden von einer allgemeinen Disposition der Besserwisserei o.ä. abhängt, nicht ohne weiteres aufgeben werden sollte.

Grundsätzlich lässt sich auch diskutieren, ob Lazarsfeld mit seinem Problem der Offensichtlichkeit überhaupt den Nagel auf den Kopf getroffen hat und ob es nicht vielmehr um die Frage von Trivialität geht. Ob die Personen das Umfrageergebnis „interessant“ finden, ob sie „das so vermutet hätten“, erscheint letztlich nicht entscheidend: Die Personen können ja durchaus ein Ergebnis richtig vorhersagen, es dann aber trotzdem nicht trivial finden, weil das Ergebnis den Status einer Meinung von dem einer Hypothese zu dem einer bestätigten Hypothese, also salopp gesprochen von Vermutung zu Wissen, verändert hat. Aus wissenschaftlicher Sicht jedenfalls ist es nicht entscheidend, dass ein Ergebnis überraschend ist, sondern dass es ein Datum – und keine noch so überzeugende Vermutung – ist, das auch anders hätte ausfallen können. Armstrong (2002: 5) bemerkt hierzu richtig, dass „findings need not be surprising to be important“. In der Tat sind die meisten wissenschaftlichen Befunde eher nicht überraschend. So zeigen Gordon u.a. (1978) z.B., dass Nichtpsychologen drei Viertel der Ergebnisse von 61 publizierten Forschungsstudien auf Grund der Beschreibung der Studie richtig voraussagen. Es bleibt also weiteren Studien überlassen, wie man durch eine geeignete Technik die scheinbare Trivialität von Umfragebefunden direkt reduzieren kann. Möglicherweise ist der Eindruck der Nicht-Offensichtlichkeit beim Nichtwissenschaftler aber leichter zu erzeugen als der von Nicht-Trivialität.

Problematisch bleibt bei dieser Untersuchung, dass die präsentierten Umfragebefunde nur fiktiv waren und somit die Frage danach, ob man das alles hätte vorhersagen können, keine objektive Bewertungsgrundlage hat. Ebenso problematisch ist es, dass die meisten Umfragethemen für die Vpn trotz aller Bemühungen im Pretest natürlich keine besondere Relevanz hatten. Trotzdem meinten die Vpn, die eigene Prognosen abgegeben haben und/oder die die Prognosen anderer gesehen haben, nach dem Experiment verstärkt, dass psychologische Forschung ganz allgemein sinnvoll ist – vielleicht auch deshalb, weil ihre Ergebnisse doch nicht so leicht vorhersagbar sind.

Bedenkenswert ist für das obige Experiment auch noch, dass Fehlbewertungen für die Vpn ohne offensichtliche negative Folgen blieben. Das ist im oben beschriebenen Kontext von Mitarbeiterbefragungen ganz anders. Dort sind die Daten für die Führungskräfte potentiell sehr bedeutsam: Fehlbewertungen können hier dazu führen, dass Geld verloren wird oder dass zumindest Chancen vertan werden. Wir erwarten daher, dass sich die Prognosen in einem relevanten Umfragekontext mit schwerwiegenden Folgen für die Rezipienten deutlicher auswirken sollten. Weitere Untersuchungen dazu werden wir noch anstellen.

Korrespondenzadresse

Prof. Dr. Ingwer Borg
ZUMA
Postfach 12 21 55
68 072 Mannheim
email: borg@zuma-mannheim.de

Carola Hillenbrand
Henley Management College, Greenlands
Henley-on-Thames
Oxfordshire, RG9 3AU, England

Literaturhinweise

- Armstrong, J.S., 2002: Discovery and communication of important marketing findings: Evidence and proposals. *Journal of Business Research*, 5775: 1-16.
- Baratz, D., 1983: How justified is the „obvious“ reaction? *Dissertation Abstracts International* 44/02B, 644B. Zitiert nach D.G. Myers 1995: *Social Psychology*. New York: MacGraw-Hill.
- Borg, I., 2002: What makes follow-up processes successful? Vortragspapier für die IT Survey Group. Boston, USA. April.
- Borg, I., 2003: *Führungsinstrument Mitarbeiterbefragung*. Göttingen: Verlag für angewandte Psychologie. 3. Auflage.
- Borg, I./Groenen, P.J.F., 1997: *Modern multidimensional scaling*. New York: Springer.
- Fiedler, K., 1993: Kognitive Täuschungen: Faszination eines modernen Forschungsprogramms S. 7-12 in: W. Hell (Hrsg.), *Kognitive Täuschungen: Fehlleistungen und Mechanismen des Urteilens, Denkens und Erinnerns*. Heidelberg: Spektrum.
- Fischhoff, B., 1977: Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance* 3: 349-358.
- Gordon, M.E./Kleinman, L.S./Hanie, C.A., 1978: Industrial-organizational psychology: open they ears o house of Israel. *American Psychologist* 33: 893-905.
- Lazarsfeld, P.F., 1949: The American soldier – an expository review. *Public Opinion Quarterly* 13: 377-404.
- Meyer, U., 1999: Entwicklung eines Fragebogens zum Konstrukt Narzismus. Forschungsbericht. Fachbereich 06 Psychologie, Uni Gießen.
- Meyer, U., 2000: Quantitative und qualitative Differenzierung narzistischer Persönlichkeitsmerkmale. In: C. Tarnai/J. Reinecke (Hrsg.), *Angewandte Klassifikationsanalyse*. Münster: Waxmann.
- Meyers, D.G., 1994: *Exploring social psychology*. New York: MacGraw-Hill.

Pohl, R., 1992: Der Rückschau-Fehler: Systematische Verfälschung der Erinnerung bei Experten und Novizen. *Kognitionswissenschaft* 3: 38-44.

Sanna, L.J./Schwarz, N./Stocker, S.L., 2002: When debiasing backfires: Accessible content and accessibility experiences in debiasing hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28: 497-502.

Stahlberg, D./Eller, F./Rohmann, A./Frey, D., 1995: Der Knew-it-all-along-Effekt in Urteilsituationen von hoher und geringer Selbstwertrelevanz. *Zeitschrift für Sozialpsychologie* 24: 94-102.

COMPUTERUNTERSTÜTZTES PRETESTING VON CATI-FRAGEBÖGEN: DAS CAPTIQ-VERFAHREN

FRANK FAULBAUM, MARC DEUTSCHMANN & MARTIN KLEUDGEN

Der vorliegende Beitrag stellt ein am Sozialwissenschaftlichen Umfragezentrum der Universität Duisburg-Essen entwickeltes computerunterstütztes, auf dem Prinzip des Behavior-Coding basierendes Pretestverfahren für CATI-Instrumente unter Feldbedingungen vor. Feldpretests (auch: Beobachtungs- oder Standardpretests) für CATI-Instrumente erfordern die Evaluation von CATI-Fragebögen im Rahmen einer computerunterstützten telefonischen Pretesterhebung. Die Registrierung der während der Pretestinterviews auftretenden Probleme mit einzelnen Fragen kann traditionell von den Pretestinterviewern entweder retrospektiv nach der Beendigung des telefonischen Pretests durch Ausfüllen eines speziellen Beobachtungsinstruments oder durch Paper- und Pencil-Notierung der Probleme während einer Frage-Antwort-Episode vollzogen werden. Weder von der retrospektiven Methode noch von der Methode des gleichzeitigen schriftlichen Eintrags sind, insbesondere bei schwierigen Erhebungsinstrumenten reliable Ergebnisse zu erwarten. Im zuletzt genannten Fall muss davon ausgegangen werden, dass der Interviewer in der Regel so sehr auf das eigentliche Interview und die computerunterstützte Eingabe der Antworten konzentriert ist, dass er mit der zusätzlichen Aufgabe, Probleme schriftlich festzuhalten, überfordert wäre. Das hier vorgestellte Verfahren versucht, diese Defizite zu umgehen und zugleich die Vorteile computerunterstützter Telefonumfragen für Pretests zu nutzen. Ein Vorteil besteht etwa in der Möglichkeit, größere, nach dem Zufallsverfahren von Gabler und Häder (1997) gezogene Preteststichproben zu verwenden, um anspruchsvollere statistische Verfahren schon in der Pretestphase einzusetzen. Ein weiterer Vorteil liegt in der raschen Verfügbarkeit eines Pretest-Datensatzes. Das hier beschriebene CAPTIQ-Verfahren (CAPTIQ: Computer Assisted Pretesting of Telephone Interview Questionnaires) erlaubt die sofortige computerunterstützte Codierung der Problemarten und liefert einen Datensatz, auf dessen Grundlage die Pretestergebnisse in ihrer longitudinalen Qualität sichtbar gemacht werden können. Dies geschieht mit Hilfe einer grafischen Darstellung, die als IPG (Interview Process Graph) bezeichnet wird.

Ähnlich wie in einem Elektrokardiogramm lassen sich Probleme mit Antwortskalen, Antworttendenzen, Lernprozesse, etc, insbesondere bei längeren Itembatterien identifizieren.

Observational pretesting or standard pretesting of CATI-questionnaires is not unproblematic because the recording of observed respondent behavior has either to be carried out during the interview itself or after completion of the interview, prepared by filling out observational forms. Recording during the interview often places a heavy additional burden on the interviewer above and beyond conducting the interview properly. Recording after the interview presents a challenge to reliability. In this paper we present a method for Computer Assisted Pretesting of Telephone Interview Questionnaires (CAPTIQ) which allows respondents' behavior to be coded during the interview without burdening the interviewer too much. The interviewers are able to code without interrupting the flow of the interview itself. The pretest data for each question and each respondent collected by CAPTIQ may be seen as longitudinal data which can be represented by a graph called IPG (Interview Process Graph). The IPG, rather like an electrocardiogram, reveals any problem zones occurring during the interview. As a result, problems concerning response scales, information collected on the learning processes initialized by respondents while processing item batteries. Comprehension difficulties related to question wording or other factors also manifest themselves in oscillations of the IPG. The paper describes the CAPTIQ-Method and presents an illustration of the IPG by evaluating a CATI-Questionnaire used for a nationwide survey about health and media use.

1. Zielsetzungen

Intensive Forschungsanstrengungen im Bereich der Optimierung von Pretestverfahren (vgl. z.B. Exposito/Rothgeb 1997; Presser/Blair 1994; Prüfer/Rexroth 1996) haben bisher noch nicht zu einem praktisch einsetzbaren Verfahren zur Optimierung von CATI-Fragebögen unter Feldbedingungen geführt. Der vorliegende Beitrag berichtet über ein computerunterstütztes Pretestverfahren, das diese Lücke schließen soll. Im Unterschied zu den kognitiven Laborverfahren wie z.B. „*think aloud*“, „*paraphrasing*“, „*probing*“, etc. verlassen sich reine Beobachtungspretests nur auf die passive Beobachtung des Befragtenverhaltens. Idealerweise sollten den Beobachtungsverfahren andere Verfahren wie z.B. die Anwendung von Appraisal-Systemen (vgl. Lessler/Willis 1999) zur Bewertung der Qualität von Fragen anhand bestimmter Kriterien oder kognitive Pretests zum Fragenverständnis vorausgegangen sein.

Das hier vorgestellte, als *CAPTIQ – Verfahren* (Computer Assisted Pretesting of Telephone Interview Questionnaires) bezeichnete Pretestverfahren für CATI-Instrumente

stellt einen ersten Vorschlag dar, eine Codierung des Befragtenverhaltens *computerunterstützt* an einer *großen Zufallsstichprobe der Zielpopulation* durchzuführen und dabei gleichzeitig in den natürlichen Ablauf eines CATI-Interviews zu integrieren, ohne dass dieser Ablauf durch die Intervieweraktivität der Codierung gestört würde (vgl. auch Deutschmann/Faulbaum/Kleudgen 2003). Insofern handelt es sich über den klassischen Beobachtungspretest hinaus, bei dem der Interviewer die Reaktionen des Befragten nur notiert, um eine Variante des Behavior Coding. In seiner klassischen Form erfolgt die Codierung *durch den Forscher nach dem Interview* auf der Grundlage von Tonbandaufnahmen. Rein technisch betrachtet, wäre eine akustische Aufzeichnung auch bei CATI-Interviews möglich. Da hierfür jedoch das vorherige Einverständnis des Befragten eingeholt werden muss, sind systematische Auswahlwirkungen zu erwarten, die die Qualität der zufällig gezogenen Preteststichprobe gefährden könnten. Darüber hinaus stellt das Wissen von der Aufzeichnung des Interviews beim Befragten eine entscheidende Abweichung von den realen Feldbedingungen dar, die sich möglicherweise erheblich auf die Reaktionen des Befragten auswirken kann.

Da die computerunterstützte Verhaltenscodierung nicht durch den Forscher selbst, sondern durch Interviewer während eines normalen CATI-Interviews erfolgen muss, das sich im Vergleich zu anderen Erhebungsmethoden durch einen besonders starken Zeitdruck auszeichnet, muss das Codiersystem sehr einfach gehalten werden und für den Interviewer leicht handhabbar sein. Dennoch bleibt die Aufgabe des simultanen Interviewens und Codierens eine gewisse Belastung für den Interviewer, die nur durch eine entsprechende Schulung und die Auswahl kompetenter und erfahrener Interviewer reduziert werden kann.

Die Begleitforschung im Zusammenhang mit dem hier vorgestellten CAPTIQ-Verfahren ist noch nicht abgeschlossen. Insbesondere fehlen Studien zur Intercoder-Reliabilität, die zwangsläufig eine akustische Aufzeichnung erfordern würden.

2. Der Prozess der Codierung

Die im Folgenden dargestellten Codierregeln sind aus anderen Systemen des Behavior-Coding abgeleitet (vgl. Fowler/Cannell 1996; Morton-Williams 1979; Oksenberg/Cannell/Kalton 1991; Prüfer/Rexroth 1985) und an die Besonderheiten der computerunterstützten Telefonumfrage angepasst worden. Um eine simultane Codierung während des computerunterstützten Interviews zu ermöglichen, musste ein Weg gefunden werden, um das Codiersystem in die CATI-Software so zu integrieren, dass der Interviewer die Codierung während des Interviews bewerkstelligen kann. Dies geschieht dadurch, dass bestimmte Funktionstasten für spezifische Typen des Befragtenverhaltens reserviert werden.

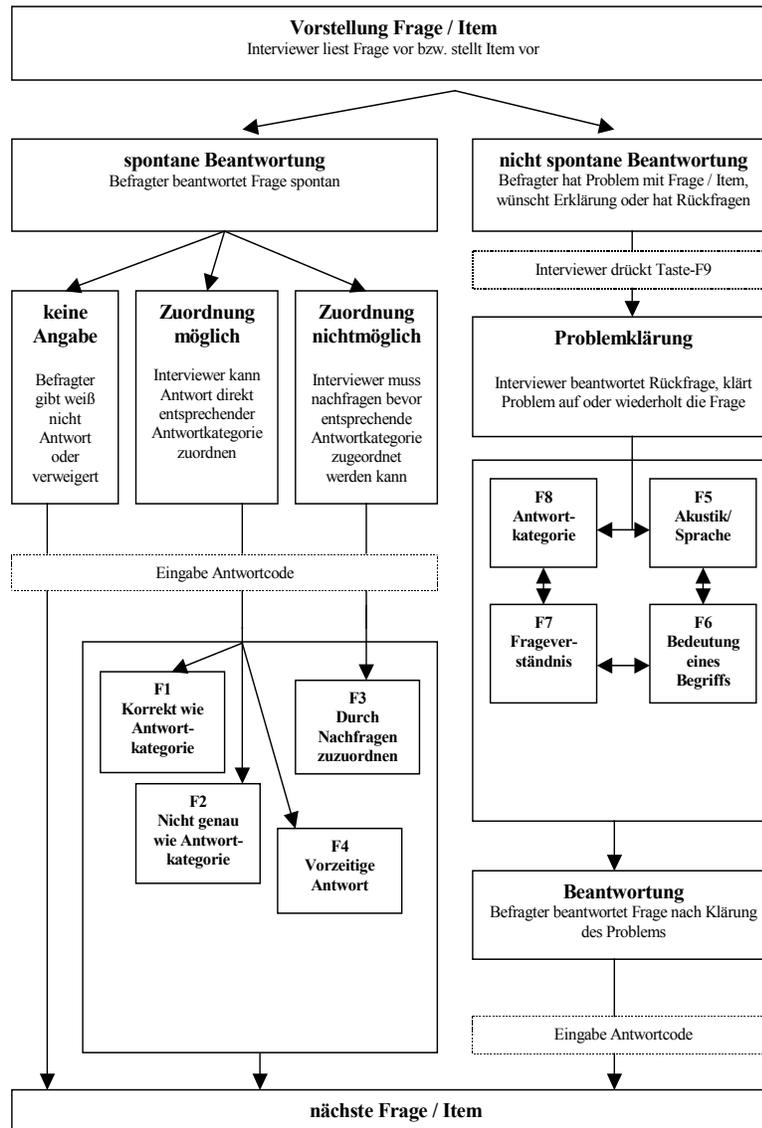
Die grundlegende Idee der Codierung des Befragtenverhaltens lässt sich durch das illustrieren, was Zouwen, Dijkstra and Ongena (2000) eine “paradigmatische Frage-Antwort-Folge“ nennen. In einer paradigmatischen, idealen und unproblematischen Frage-Antwort-Folge liest der Interviewer korrekt die Fragen vor, stellt die Antwortkategorien vor und der Befragte gibt Antworten, die der Interviewer einer von mehreren möglichen Antwortkategorien zuordnen kann; d.h. der Befragte gibt nur *adäquate* Antworten. Es ist wohlbekannt, dass die Einstufung einer Antwort als adäquat nicht ausreicht, um die Bedeutung der Befragtenreaktion zu verstehen. So antworten etwa Befragte u.U. auch dann adäquat, wenn sie die Frage nicht verstanden haben. Leider gibt es im Rahmen eines Beobachtungspretests keine Möglichkeit zu entscheiden, wann dies der Fall ist. Ziel der hier beabsichtigten Codierung ist die Klassifikation der Antworten als *adäquat* oder *inadäquat* und die Identifikation *ausgewählter Typen nicht adäquater* Antworten. Da nur eine Codierung des Befragtenverhaltens und nicht des Interviewerverhaltens möglich ist, kann nicht entschieden werden, ob inadäquates Befragtenverhalten durch inadäquates Interviewerverhalten verursacht wurde. Die Wahrscheinlichkeit der zuletzt genannten Möglichkeit kann durch extensives Interviewertraining weiter reduziert werden. Überdies stellt dies bei großer Preteststichprobe und vielen Pretestinterviewern kein ernsthaftes Problem dar, da die Bedeutung systematischer Einflüsse durch Interviewer in der statistischen Analyse kontrolliert werden kann. Insbesondere können durch eine interviewerspezifische Auswertung (siehe unten) relative Schwächen der Interviewer aufgedeckt werden.

Abbildung 1 beschreibt Struktur und Prozess der Codierung sowie die während des Codierprozesses verwendeten Funktionstasten. Selbstverständlich kann die Zuordnung der Reaktionen des Befragten zu Funktionstasten auch anders programmiert werden. Das hier zugrundegelegte Codiersystem unterscheidet grundsätzlich zwischen zwei Arten von Befragtenverhalten:

- Spontane Beantwortung der Frage
- Nicht-spontane Beantwortung der Frage

Mit dem Begriff der *spontanen Beantwortung* soll ausgedrückt werden, dass der Befragte den Versuch unternimmt, in seiner ersten Reaktion auf die Frage eine adäquate Antwort zu geben. Diese Antwort kann entweder vollständig adäquat in dem Sinne sein, dass die Antwort durch den Interviewer einer der zulässigen Antwortkategorien zugeordnet werden kann oder sie ist insofern inadäquat, als der Interviewer eine Zuordnung nicht ohne Nachfragen zuordnen kann. Im Einzelnen können folgende *Varianten einer spontanen Beantwortung* unterschieden werden:

Abbildung 1: Ablauf der Codierung



- **Antwort entspricht genau einer Antwortvorgabe:** Befragter antwortet genau wie in der Antwortvorgabe vorgesehen, verwendet die gleichen Wörter wie in der Antwortvorgabe, Antwort ist problemlos zuzuordnen (Interviewer gibt den Antwortcode ein und codiert anschließend durch Drücken der Funktionstaste **F1**, dass die Frage problemlos zuzuordnen war).
- **Antwort entspricht nicht ganz genau einer Antwortvorgabe:** Befragter antwortet nicht genau wie in der Antwortvorgabe vorgesehen, verwendet andere oder ähnliche Wörter wie in der Antwortvorgabe, Antwort ist jedoch ohne weitere Nachfragen noch gut zuzuordnen (Interviewer gibt den Antwortcode ein und codiert anschließend durch Drücken der Funktionstaste **F2**, dass die Antwort nicht ganz genau der Antwortvorgabe entsprach).
- **Durch Nachfragen zuzuordnen:** Befragter antwortet zwar direkt, muss aber noch einmal gefragt werden, wo seine Antwort zugeordnet werden kann, Richtung der Antwort ist klar – die genaue Ausprägung aber noch nicht (Interviewer fragt nach, gibt die Antwort ein und codiert anschließend durch Drücken der Funktionstaste **F3**, dass eine Zuordnung nur durch Nachfragen möglich war).
- **Vorzeitige Antwort:** Befragter antwortet zu früh, kennt eigentlich noch nicht die gesamte Frage / Aussage und antwortet einfach während des Vorlesens der Frage/des Items (Interviewer gibt Antwort ein und codiert anschließend durch Drücken der Funktionstaste **F4**, dass die Antwort vorzeitig erfolgte). **Verweigerung oder „Weiß-nicht“** (direkte Eingabe der dafür vorgesehenen Antwortcodes Standardmode ohne anschließendes Drücken einer Funktionstaste).

Eine *nicht-spontane Beantwortung* liegt dann vor, wenn der Befragte in einer ersten Reaktion vom Interviewer eine Klärung verlangt, eher er eine Antwort abgeben kann oder sich zu einer „weiß nicht“-Antwort oder einer Verweigerung entschließt. In diesem Fall drückt der Interviewer in jedem Fall erst einmal die Funktionstaste F9, um damit zu codieren, dass der Befragte mit der Frage/dem Item Probleme hatte. Der Interviewer klärt dann in Übereinstimmung mit den Regeln des standardisierten Interviews, gibt nach Klärung die Antwort oder „weiß nicht“ oder „verweigert“ ein und codiert anschließend durch Drücken der Funktionstasten F5, F6, F7 oder F8, um was für eine Problemart es sich handelte, wobei folgende Problemarten unterschieden werden:

- **Akustik / Sprache:** Befragter hat den vorgelesenen Frage- bzw. Itemtext nicht richtig gehört, versteht die Sprache nicht gut, die Telefonverbindung ist schlecht, Geräusche in der Leitung (Interviewer drückt **F5**).
- **Bedeutung eines Begriffs:** Komplizierter Begriff wird nicht verstanden, Bedeutung von Fremdwörtern werden nicht genau verstanden, Befragter kennt einen Begriff oder ein bestimmtes Wort nicht (Interviewer drückt Funktionstaste **F6**).

- **Frageverständnis:** Befragter versteht die Frage/Item Aussage in ihrer Bedeutung nicht richtig, versteht den Sinn dieser Frage/Item nicht, versteht nicht, warum diese Frage gestellt wurde (Interviewer drückt Funktionstaste **F7**).
- **Antwortkategorien:** Befragter weiß nicht mehr, wie er zu antworten hat, hat die Antwortalternativen vergessen, zu komplizierte Antwortskala (Interviewer drückt Funktionstaste **F8**).

Diese Problemarten sind im Prinzip noch weiter differenzierbar, wobei bedacht werden muss, dass jede weitere Verfeinerung zusätzliche Anforderungen an das Training der Pretestinterviewer bedeutet. Dennoch wird gegenwärtig an einer Lösung des Problems gearbeitet, wie das Codiersystem weiter verfeinert werden kann, ohne dass das Interviewertraining zu anspruchsvoll wird. Insbesondere ist zu überlegen, ob die oben aufgeführten Problemarten in sich nicht noch zu heterogen sind und weiter zerlegt werden sollten.

Wie in jedem anderen System des Behavior-Coding, kann man auch in diesem Fall allein auf Basis der Pretestdaten nicht immer eindeutig entscheiden, ob die beobachteten Abweichungen vom idealen Frage-Antwortverlauf auf den Befragten, die Frage oder die Antwortkategorien zurückzuführen ist. Hierzu wäre eine detailliertere Betrachtung der entsprechenden Interviewsequenzen erforderlich. Durch das CAPTIQ-Verfahren wird aber eine Aufdeckung der problematischen Stellen im Interview ermöglicht. Da das Interviewerverhalten nicht beobachtet wird, kennen wir ebenfalls nicht das Ausmaß, in dem das beobachtete Befragtenverhalten vom Interviewer selbst beeinflusst wird. Wenn etwa Äußerungen beobachtet werden, die in die Problemart „Akustik/Sprache“ fallen, so kann das mangelnde akustische Verstehen etwa durch eine undeutliche oder zu leise Sprache des Interviewers bei gutem Hörvermögen des Befragten oder an einem schlechten Hörvermögen bzw. der akustischen Inkompetenz des Befragten bei deutlichem Sprechen des Interviewers liegen.

3. Visualisierung und Analyse der Pretestergebnisse: Der Interview Process Graph (IPG)

Wie bereits oben erwähnt, bietet das CAPTIQ-Verfahren die Möglichkeit, größere, zufällig gezogene Preteststichproben einzusetzen, was nicht nur den Einsatz komplexerer statistischer Verfahren wie verschiedener Varianten der Faktorenanalyse, der multidimensionalen Skalierung oder der Analyse von Strukturgleichungsmodellen erlaubt und damit den Einsatz fortgeschrittener Verfahren der Itemanalyse, der Überprüfung theoretischer Validität, Konstruktvalidität und Reliabilität schon in der Pretestphase ermöglicht. In seiner Effizienz nicht zu unterschätzen ist vielmehr die Anwendung von Verfahren der

Visualisierung, die einen Überblick über die Probleme mit Fragen/Items im vollständigen Ablauf des Interviews, also im Längsschnitt erlauben.

So lassen sich etwa die Häufigkeiten und Prozentwerte der vergebenen Codes für jede einzelne Frage bzw. jedes Item in der originalen Reihenfolge der Fragen und Items im Fragebogen abtragen. Es entsteht ein Verlaufsdiagramm, das für jede Frage und jedes Item die Häufigkeit des Auftretens der entsprechenden Probleme sichtbar macht. Diese Form der Visualisierung der Codeverteilungen im Zeitverlauf des Interviews nennen wir *Interview Process Graph (IPG)*, wobei der Begriff unabhängig von der Art des oder der dargestellten Codes definiert ist. Wir können so einen IPG für Verständniscodes, für Verweigerungen, etc. unterscheiden. In einem IPG können auch Eigenschaften verschiedener Codeverteilungen, z.B. Verständnisprobleme *und* Verweigerungen, gemeinsam abgetragen werden. Ferner lassen sich verschiedene Codes zu allgemeineren Problemarten zusammenfassen. So kann man etwa auf der Basis der vergebenen Codes die Verteilungen der in- adäquaten Reaktionen im Zeitverlauf darstellen.

IPGs erlauben die Identifikation möglicher Problemzonen im Interviewverlauf und die Analyse von Problemen mit Fragen/Items im Kontext von Nachbarfragen/Items, was z.B. bei längeren Itembatterien von Interesse ist. Sie erlauben auch die Visualisierung von Lern- und Anpassungsprozessen, die im Interviewverlauf auftreten. So kann man etwa erkennen, wie schnell Befragte lernen, eine bestimmte Form von Antwortskala zu handhaben.

Das CAPTIQ-Verfahren wurde bisher in vier Umfragen des Sozialwissenschaftlichen Umfragezentrums der Universität Duisburg-Essen erprobt. Als Beispiel sei eine Umfrage zum Thema „Medien und Gesundheit“ herausgegriffen, die im Rahmen einer Studie des Deutschen Diabetes-Forschungsinstituts (DDFI) an der Heinrich-Heine-Universität Düsseldorf durchgeführt wurde.¹ Es ging in dieser bundesweiten Studie um das Mediennutzungsverhalten der Bevölkerung bei der Gewinnung gesundheitsbezogener Informationen. Das Instrument beinhaltete sowohl einfache Ja/Nein-Fragen zu gesundheitlichen Beschwerden und Krankheiten als auch umfangreichere Itembatterien mit Zustimmungsskalen zum Thema Gesundheit, Fragen zur Informiertheit über bestimmte Erkrankungen sowie Fragen zur Mediennutzung allgemein. Insgesamt bezogen sich 124 Fragen/Items auf diese Inhalte, hinzu kamen Screening-Fragen sowie demographische Fragen, die allerdings nicht in den Pretest mit einbezogen wurden.

Für den CATI-Pretest wurden sechs besonders erfahrene Interviewer ausgewählt. Mit ihnen und einigen Supervisoren wurden dann die angesprochenen eingehenden

¹ Wir danken an dieser Stelle besonders Prof. Dr. Werner Scherbaum für die freundliche Kooperation in diesem Projekt.

Pretestschulungen vorgenommen. Insgesamt wurden 100 CATI-Pretestinterviews mit dem CAPTIQ-Verfahren durchgeführt, die nach dem Verfahren von Gabler/Häder (1997, 1998, 1999) rekrutiert wurden.

Nach Anwendung des Behavior-Codings erschien es sinnvoll, die Fülle der erhobenen Informationen zum Befragtenverhalten zu den folgenden Kategorien zusammenzufassen:

- Spontane, adäquate oder nahezu adäquate Beantwortung (Codes: F1, F2, F4)
- spontane, inadäquate Beantwortung (Code: F3)
- nicht-spontane Beantwortung aufgrund eines Problems (Codes: F5, F6, F7, F8).

Mit dieser Klassifikation kann der gesamte Interviewverlauf im Interview-Process-Graph (IPG) gut und übersichtlich visualisiert werden. Es zeigt - ähnlich wie ein Elektrokardiogramm (EKG) die Ströme des menschlichen Herzens - die Problemzonen im Verlauf eines gesamten Telefoninterviews. Einfache und schwierige Antwortskalen können so genau identifiziert werden wie auch einsetzende Lernprozesse seitens des Befragten bei umfassenden Itembatterien. Verständnisprobleme und Schwächen einzelner Fragen machen sich durch extreme Ausschläge im IPG bemerkbar.

Der in Abbildung 2 dargestellte IPG enthält die Prozentwerte für alle 124 Fragen mit inhaltlichem Bezug, mit denen die oben dargestellten Verhaltenskategorien bei den einzelnen Fragen/Items auftraten. Trotz ihrer eingeschränkten Lesbarkeit wurde die Grafik dargestellt, um einen visuellen Gesamteindruck des IPG zu ermöglichen.

Interessant scheint auch ein weiteres Phänomen, das bei dieser Itembatterie besonders gut zu sehen ist, aber auch bei anderen Itemfolgen auftaucht. Die erstgenannten Items zeigen gegenüber den letztgenannten schlechtere Werte im Antwort- und Befragtenverhalten. Dies könnte damit begründet werden, dass der Befragte seine zu erfüllende Aufgabe entweder nach jedem Item besser beherrscht oder der Befragte in bestimmte Antworttendenzen verfällt. Die Vorlage des ersten Items FR18_1 verursacht noch bei 17% der Befragten Probleme, so dass eine spontane Zustimmung oder Ablehnung ausbleibt. Diese Probleme sind vor allem auf das Item- bzw. Frageverständnis (7%) und Probleme mit der Antwortkategorie (6%) zurückzuführen. Bei 4% der Befragten musste das Item lediglich wiederholt vorgelesen werden. Die darauffolgenden Items verursachten deutlich weniger Probleme. Die entsprechenden Prozentwerte aus dem IPG sind in Tabelle 1 noch einmal zusammengefasst. Innerhalb dieser und anderer Itembatterien könnte also die steigende Anzahl spontan gegebener adäquater Antworten und die abnehmende Anzahl inadäquater und nicht-spontan gegebener Antworten mit einem einsetzendem Lernprozess des Befragten erklärt werden.

Abbildung 2: Beispiel eines Interview-Process-Graph (IPG)

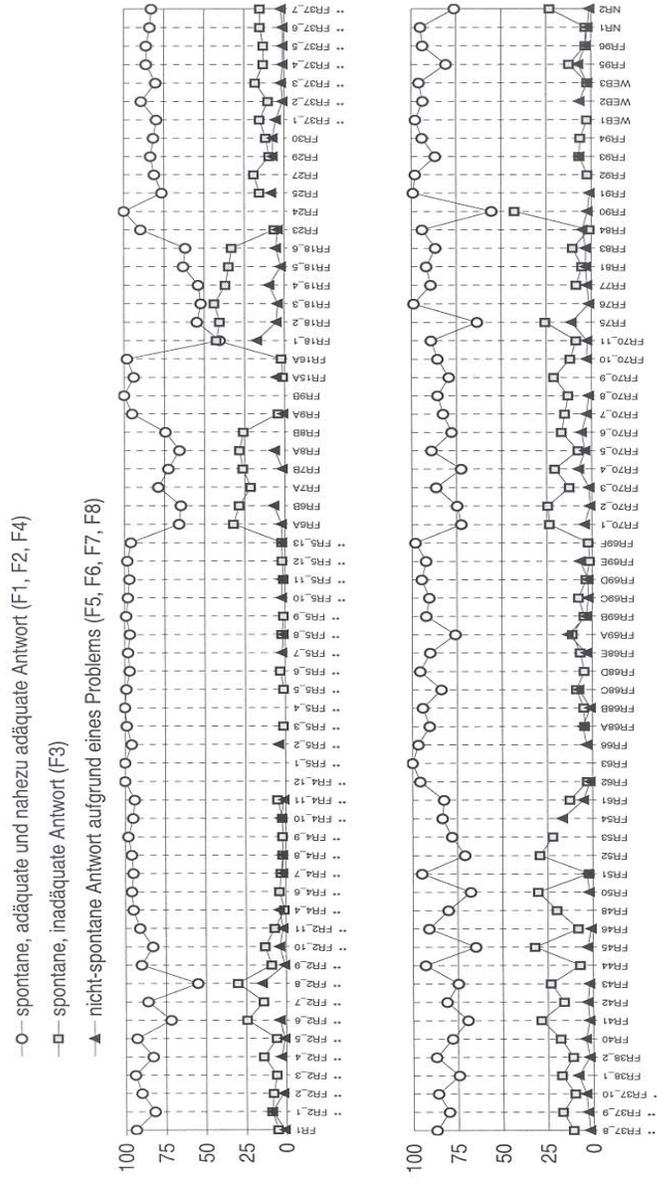


Tabelle 1: Anteilswerte adäquater und inadäquater Antworten (N=100)

	spontane, adäquate und nahezu adäquate Antwort (F1, F2, F4)	spontane, inadäquate Antwort (F3)	nicht-spontane Antwort aufgrund eines Problems (F5, F6, F7, F8)
FR18_1	40,4	42,4	17,2
FR18_2	54,5	40,4	5,1
FR18_3	52,1	43,8	4,2
FR18_4	53,7	36,8	9,5
FR18_5	63,0	34,8	2,2
FR18_6	61,7	33,0	5,3

4. Befragten- und interviewerspezifische Analysen

4.1 Befragten-spezifische Analyse

Bisher wurde die itemspezifische Analyse des Antwort- und Befragtenverhaltens unter dem Aspekt der qualitativen Bewertung des Instruments geschildert. Aufgrund der Möglichkeit, bei dem CAPTIQ-Verfahren relativ große Stichprobenumfänge einsetzen zu können, scheint die Methode aber auch für befragten- und interviewerspezifische Analysen des Befragtenverhaltens geeignet zu sein. So lassen sich etwa Fragen beantworten wie z.B.: Gibt es Befragtengruppen, die besonders viele Probleme bei bestimmten Fragen oder Formulierungen haben? Oder: Welche Merkmale der Befragten haben einen Einfluss auf ihr Antwortverhalten im Interview?

Als Beispiel greifen wir einige demographische Merkmale der Befragten (Geschlecht, Alter, Bildungsgrad) heraus und setzen sie in Beziehung zum Antwortverhalten. Tabelle 2 zeigt die Summenwerte der zusammengefassten Codes über alle Fragen/Items sowie über alle Interviewer. Eine Analyse der inadäquaten Beantwortungen ergibt durchaus sichtbare Unterschiede. Personen im Alter ab 45 Jahren und Personen mit niedrigerem Bildungsabschluss geben deutlich mehr spontane inadäquate Antworten als der Durchschnitt, Personen im Alter von 60 Jahren und darüber sowie Personen mit niedrigerem Bildungsabschluss geben auch mehr nicht-spontane inadäquate Antworten.

Dies sind zwar keine besonders überraschenden Ergebnisse, sie unterstreichen aber an dieser Stelle die Plausibilität des vorgestellten Verfahrens. Ähnliches fanden auch Prüfer und Rexroth (1985) in ihren Arbeiten über die Interaction-Coding-Technik heraus sowie

Reuband (1998) in seiner Befragung der Interviewer über der Verständnisprobleme seitens der Befragten.

Tabelle 2: Anteile demografischer Merkmale (N=100)

		spontane, adäquate und nahezu adäquate Antwort (F1, F2, F4)	spontane, inadäquate Antwort (F3)	nicht-spontane Antwort aufgrund eines Problems (F5, F6, F7, F8)
Geschlecht	männlich	86,8	9,9	3,3
	weiblich	84,2	13,0	2,8
Alter	16 - 29 Jahre	89,8	7,3	2,9
	30 - 44 Jahre	86,9	9,7	3,3
	45 - 59 Jahre	83,3	14,1	2,5
	60 Jahre und mehr	78,6	17,9	3,6
Schulabschluss	niedriger Schulabschluss	80,8	15,9	3,3
	höherer Schulabschluss	88,2	9,1	2,7
Gesamt		85,3	11,6	3,0

4.2 Interviewerspezifische Analyse

Mit umfangreicheren Preteststichproben lässt sich auch der Einfluss der Pretestinterviewer auf die Ergebnisse des Behavior-Codings überprüfen. Bereits auf der Ebene einfacher deskriptiver Analysen wird sichtbar, ob die Interviewer sich in ihrer Verhaltensklassifikation von einander unterscheiden. Da die Befragten bzw. die Telefonnummern durch das CATI-Programm nach Zufall auf die Interviewer verteilt wurden, haben wir eine Randomisierung über den Faktor „Interviewer“ vorliegen. Tabelle 3 stellt für jeden Interviewer die prozentualen Häufigkeiten der von ihm vergebenen Problemkategorien dar, und es werden durchaus Unterschiede sichtbar.

Während z.B. Interviewer BM bei durchschnittlich 6% seiner Interviews Antwortprobleme vercodete, waren es bei den Interviewern GA und ZI nur 1,6%. Auf der anderen Seite weist Interviewer ZI den höchsten Anteil an Befragten mit der Klassifikation „spontane inadäquate Antwort“ auf. Die Ergebnisse deuten darauf hin, dass ein solcher Pretest trotz intensiver Interviewerschulung nicht frei von Interviewereffekten ist. Das durch

diese Methode aufgedeckte quantitative Ausmaß dieser Effekte kann aber bei der Bewertung der Pretestergebnisse entsprechend mit berücksichtigt werden.

Tabelle 3: Codierung der Interviewer im Vergleich

	Anzahl der durchgeführten Interviews	spontane, adäquate und nahezu adäquate Antwort (F1, F2, F4)	spontane, inadäquate Antwort (F3)	nicht-spontane Antwort aufgrund eines Problems (F5, F6, F7, F8)
Interviewer: AE	12	84,3	11,5	4,2
Interviewer: BM	18	81,7	12,3	6,0
Interviewer: GA	20	92,2	6,2	1,6
Interviewer: KA	11	86,2	11,1	2,7
Interviewer: SC	12	86,1	10,5	3,3
Interviewer: ZI	27	82,9	15,5	1,6

5. Schlussbemerkung

Das computerunterstützte Behavior-Coding unter Feldbedingungen mit großen Zufalls-Preteststichproben nutzt die spezifischen Vorteile von CATI-Umfragen und ermöglicht Auswertungen, die bei den üblichen persönlich/mündlichen Pretests nicht oder nur mit sehr großem Aufwand möglich sind. Das Verfahren arbeitet auf seinem gegenwärtigen Entwicklungsstand noch recht grob und ist daher nur beschränkt in der Lage, konkrete Verbesserungsvorschläge für bestimmte Fragen und Itembatterien anzubringen. An einer Optimierung der Methode in dieser Richtung wird gegenwärtig gearbeitet. Immerhin ist es in der Lage, nicht nur einen Eindruck des Umfangs bestimmter Problemarten zu vermitteln, sondern auch detailliert bestimmte Problemzonen des Erhebungsinstruments im Längsschnitt aufzuzeigen. Die so identifizierten Bereiche können dann in einem weiteren Schritt mithilfe kognitiver Laborverfahren untersucht werden, um gesicherte Hinweise auf die Art der erforderlichen Veränderungen zu bekommen.

Das CAPTIQ-Verfahren ist für den routinemäßigen Einsatz zum Pretesten von CATI-Instrumenten entwickelt worden. Die Module zur Codierung des Befragtenverhaltens lassen sich bei Verwendung der selben Software und mutmaßlich auch in alternativer Software im Prinzip für jeden CATI-Fragebogen implementieren. Damit ist es möglich, mit begrenztem Aufwand und in relativ kurzer Zeit einen Fragebogen unmittelbar vor der Hauptfeldzeit zu testen. Falls danach keine Änderungen an den Fragen nötig sind, lassen sich die erhobenen Pretestdaten ohne weiteres für die Haupterhebung nutzen, da der Pretest unter realen Feldbedingungen stattfindet und für die Befragten nicht wahrnehmbar ist.

Korrespondenzadresse

Prof. Dr. Frank Faulbaum
Lehrstuhl für Sozialwissenschaftliche Methoden/Empirische Sozialforschung
Sozialwissenschaftliches Umfragezentrum
Universität Duisburg-Essen, Standort Duisburg
Lotharstraße 65
47048 Duisburg
email: faulbaum@uni-duisburg.de

Literatur

Deuschmann, M./Faulbaum, F./Kleudgen, M., 2003: Computer Assisted Pretesting of Telephone Interview Questionnaires (CAPTIQ). In: Proceedings of the American Statistical Association, Survey Research Section, New York: ASA.

Exposito, J. L./Rothgeb, J. M., 1997: Evaluating Survey Data: Making the Transition from Pretesting to Quality Assessment. In: Lyberg, L. et al. (eds.), Survey Measurement and Process Quality. New York: Wiley.

Fowler, F.J./Cannell, C.F., 1996: Using Behavioral Coding to Identify Cognitive Problems with Survey Questions. S. 15-36 in: Schwarz, N./Sudman, S. (Hrsg.), Answering Questions: Methodology for Determining Cognitive and Communicative Process in Survey Research. San Francisco, Jossey-Bass.

Gabler, S./Häder, S., 1997: Überlegungen zu einem Stichprobendesign für Deutschland. ZUMA-Nachrichten 41: 7-18.

Gabler, S./Häder, S., 1998: Probleme bei der Anwendung von RLD-Verfahren. S. 58-68 in: Gabler, S./Häder, S./Hoffmeyer-Zlotnik, J. (Hrsg.), Telefonstichproben in Deutschland. Opladen: Westdeutscher Verlag.

Gabler, S./Häder, S., 1999: Erfahrungen beim Aufbau eines Auswahlrahmens für Telefonstichproben in Deutschland. ZUMA-Nachrichten 44: 45-61.

- Gallhofer, I. N./Saris, W.E., 2000: Formulierung und Klassifikation von Fragen, in: ZUMA-Nachrichten 46: 43-72.
- Morton-Williams, J., 1979: The Use of "Verbal Interaction Coding" Evaluating a Questionnaire. *Quality and Quantity* 13: 59-75.
- Oksenberg, L./Cannell, Ch./Kalton, G., 1991: New Strategies for Pretesting Survey Questions. *Journal of Official Statistics* 7: 349-365.
- Presser, S./Blair, J., 1994: Survey Pretesting : Do Different Methods Produce Different Results? *Sociological Methodology*: 73-104.
- Prüfer, P./Rexroth, M., 1985: Zur Anwendung der Interaction-Coding-Technik. ZUMA-Nachrichten 17: 2-49.
- Prüfer, P./Rexroth, M., 1996: Verfahren zur Evaluation von Survey-Fragen: Ein Überblick. ZUMA-Nachrichten 39: 95-115.
- Prüfer, P./Rexroth, M., 2000: Zwei-Phasen-Pretesting. ZUMA-Arbeitsbericht 2000/8.
- Reuband, K.H., 1998: Der Interviewer in der Interaktion mit dem Befragten – Reaktionen der Befragten und Anforderungen an den Interviewer. S. 138-155 in: Statistisches Bundesamt (Hrsg.), Interviewereinsatz und –qualifikation. Band 11 der Schriftenreihe Spektrum Bundesstatistik.
- Schaeffer, N.C./Maynard, D.W., 1996: From Paradigm to Prototype and Back Again: Interactive Aspects of Cognitive Processing in Standardized Survey Interviews. S. 65-88 in: Schwarz, N./Sudman, S., (Hrsg.), Answering Questions: Methodology for Determining Cognitive and Communicative Process in Survey Research. San Francisco: Jossey-Bass.
- Van der Zouwen, J./Dijkstra, W./Ongena, Y., 2000: What Characteristics of Questions in Survey-Interviews make the Interaction Between Interviewer and Respondent 'Problematic' or Even 'Inadequate'? Department of Social Research Methodology, Vrije Universiteit, Amsterdam. Paper to presented on the Fifth International Conference on Logic and Methodology, Köln, Oktober 2000.
- Willis, G.B./Lessler, J.T., 1999: Question Appraisal System – 1999. Washington: Research Triangle Institute.

ANIMAQU – EINE COMPUTERBASIERTE BEFRAGUNG ALS MÖGLICHKEIT ZUR ERFASSUNG BESONDERER ZIELGRUPPEN. EIN BEISPIEL EINER ANWENDUNG BEI GEHÖRLOSEN MENSCHEN¹

*JOACHIM GERICH, ROLAND LEHNER
JOHANNES FELLINGER & DANIEL HOLZINGER*

Die Anwendung einer schriftlichen standardisierten Befragung ist an bestimmte Voraussetzungen der interviewten Personen gebunden. Insbesondere stellt ausreichende Lesekompetenz eine zentrale Anwendungsvoraussetzung dar. Für gehörlose Menschen - d.h. Mitglieder der Gehörlosenkultur - ist die Gebärdensprache die eindeutig bevorzugte Kommunikationsform. Die Lesekompetenz liegt im Durchschnitt unter der der Normalbevölkerung und weist eine große Variationsbreite auf. Es wird daher eine computergestützte Erhebungsform präsentiert, welche die selbstadministrierte Befragung dieser Zielgruppe ermöglicht.

Standardized paper-and-pencil interviews are not suitable for administration with every population or with every individual within a population. Respondents require to be literate, that is, able to read the language used for the interview. Severely hearing-im-

¹ In diesem Beitrag werden die Bezeichnungen „gehörlose Menschen“ oder „Menschen mit Hörbehinderung“ verwendet. Es sei jedoch angemerkt, dass damit unterschiedliche Bevölkerungsgruppen gemeint sind, und die Definition der eigenen Zugehörigkeit in der Selbstdefinition von Menschen liegt: „It is important to mention here that the degree of hearing loss is not the most important requirement for being in the core of the Deaf Community. Deaf people just identify themselves as deaf or hard of hearing, no matter what their degree of loss is“ (Kannapell 1989: 23). Nach Schein (1979) kann angenommen werden, dass jeder tausendste Mitbürger prälingual gehörlos ist. Die Prävalenzraten sind nach medizinischen Angaben in jüngster Zeit aber aufgrund der besseren frühkindlichen medizinischen Betreuung und der Früherfassung von Hörschädigungen im Abnehmen begriffen.

paired people who have been deaf from an early age generally use sign language to communicate and their level of reading ability in the national language is generally lower than of the population at large and may also vary greatly among the deaf population. The paper describes a computer-assisted form of administration which allows these deaf respondents to self-complete an interview.

1. Einleitung und Problemstellung

Die Anwendung standardisierter Erhebungsinstrumente in der empirischen Sozialforschung ist an kulturelle Rahmenbedingungen der befragten Personen geknüpft. Vorausgesetzt wird i.d.R. die Kommunikationsfähigkeit und die Bereitschaft zur Kommunikation in gesprochener und/oder schriftsprachlicher Form. Durch diese Voraussetzung resultiert ein Ausschluss unterschiedlicher Bevölkerungsgruppen als Zielgruppen dieser traditionellen Erhebungsformen. Es stellt sich die Frage, ob durch den Einsatz neuer Technologien eine Zielgruppenerweiterung bei der Anwendung standardisierter Erhebungsmethoden möglich ist.

Im Rahmen eines Forschungsprojekts zur Erhebung der Lebensqualität und der Symptombelastung bei gehörlosen Menschen in Oberösterreich² bestand die Notwendigkeit, eine gebärdensprachliche Version der Erhebungsinstrumente so zur Anwendung zu bringen, dass die Befragungsform einer ansonsten üblichen Paper-and-Pencil Befragungsmodalität sehr nahe kommt.

2. Die Zielgruppe

Der Begriff „Gehörlosigkeit“ wird hier als kulturelle Zugehörigkeit verstanden. Die Gehörlosengemeinschaft definiert sich primär über die Verwendung der Gebärdensprache, über die gemeinsame Geschichte, die Erziehung (insbesondere der Besuch von Gehörlosenschulen), über die Zugehörigkeit zu entsprechenden Gehörlosenvereinen und bevorzugte Sozialkontakte innerhalb dieser Gehörlosengemeinschaft. Die rein medizinische Sicht des Gehörlosigkeitsbegriffs im Sinne eines vollkommen fehlenden Hörvermögens spielt dabei eine untergeordnete Rolle, da Menschen mit Hörschädigungen unterschiedlichen Ausmaßes zur Gehörlosengemeinschaft zählen.

² Ein Projekt des Instituts für Sinnes- und Sprachneurologie, Krankenhaus der Barmherzigen Brüder, Projektleitung Dr. J. Fellingner, finanziert aus Mitteln des Fonds Gesundes Österreich und der Landesregierung OÖ.

Die nationalen Gebärdensprachen Gehörloser sind natürliche Sprachen, welche sowohl bezüglich ihrer strukturellen Komplexität als auch hinsichtlich ihrer Leistungsfähigkeit bei der Informationsübermittlung (Funktionalität) Lautsprachen völlig ebenbürtig sind. Die Gebärdensprachen verfügen über keine Schriftform im Sinne einer Gebrauchsschrift. Sinnerfassende Lesekompetenz ist bei einem großen Teil der gehörlosen Menschen nicht gegeben. In einer Untersuchung der Leseleistungen hochgradig hörgeschädigter Schulabgänger (ca. 16jährig) in Großbritannien (Conrad 1979) ergab sich bei 50% ein funktioneller Analphabetismus, 22% der Jugendlichen lasen auf dem Niveau von 10jährigen und nur 2,5% altersentsprechend. Dies bedeutet, dass 75% der resthörigen und 85% der völlig gehörlosen Schulabgänger keine Zeitung lesen können, wenn man als Maßstab für das Lesen einer Zeitung die Lesefähigkeit eines durchschnittlichen elfjährigen Kindes anlegt. Ähnliche Ergebnisse mit allerdings wesentlich kleineren Stichproben liegen auch für Österreich vor (Gelter 1989). Demnach können nur 10% der gehörlosen Schulabgänger eines Jahrgangs des Bundesinstituts für Gehörlosenbildung in Wien (Durchschnittsalter 14,8 Jahre) sinnverstehend lesen (d.h. zumindest auf dem Niveau eines Neunjährigen), 90% erbringen schlechtere Leseleistungen.

3. Das Erhebungsinventar

Im Rahmen des Forschungsprojektes wurden drei Erhebungsinstrumente eingesetzt. Beim ersten Erhebungsinventar handelt es sich um das WHO-Instrument zur Erfassung der Lebensqualität (The WHOQOL Group 1998). Da eine Anforderung an das Instrument in der Kompaktheit der Anwendung bestand, wurde die Kurzversion in Form des WHOQOL-BREF eingesetzt. Dieses Instrument besteht aus 26 Kernfragen, die aus dem 100 Items umfassenden WHOQOL abgeleitet wurden. Als Grundlage wurde die deutsche Version des WHOQOL-BREF verwendet (Angermeyer/Kilian/Matschinger 2000). Des Weiteren wurde eine gekürzte Form des BSI (Brief Symptom Inventory) mit insgesamt 32 Fragen verwendet (Derogatis et al. 1982). Bei beiden Instrumenten wurden ausschließlich standardisierte Items mit fünfteiligen Ratingskalen verwendet. Als drittes Instrument wurde der GHQ (General Health Questionnaire) von Goldberg (1978) mit 12 Fragen und dichotomisierten Antwortkategorien eingesetzt.

4. Konkretisierung der spezifischen Erhebungsprobleme

Die hier eingesetzte Erhebungsinstrumente sind als standardisierte schriftliche Befragungen konstruiert. Sie sind weiterhin so konstruiert, dass die Personen die Fragebögen selbstständig ausfüllen und während des Ausfüllens des Fragebogens vom Untersuchungsleiter keine Fragen mehr beantwortet werden. Für die deutschsprachige Version des WHOQOL

wird daher auch darauf hingewiesen, dass für die Anwendung des Fragebogens „ausreichende Lesefähigkeit des Probanden notwendig“ ist (Angermeyer/Kilian/Matschinger 2000: 25). Für die Population der gehörlosen Menschen ist daher sowohl für diesen speziellen Anwendungsfall, aber auch für die schriftliche standardisierte Befragung im allgemeinen davon auszugehen, dass wesentliche Voraussetzungen zur Anwendung nicht gesichert sind: „Eine schriftliche Befragung bietet sich an, wenn der schriftliche Fragebogen als Frage-Antwort-Medium kaum Schwierigkeiten begegnet. Sie eignet sich daher nicht für schreib- und denkungsgewandte Personen“ (Atteslander 1993: 164). Wenn man also davon ausgeht, dass bei der Gruppe von Menschen mit Hörbehinderung mit „Gewandtheit im Denken“, jedoch nicht immer mit Lese- und Schreibgewandtheit zu rechnen ist, so besteht die Anforderung an ein spezifisches Erhebungsinstrument darin, die zentrale Kommunikationsform der standardisierten Erhebung an die Erfordernisse dieser Gruppe anzupassen.

Eine Möglichkeit dieser Anpassung würde sich dadurch anbieten, dass das vorliegende Erhebungsinstrument – abweichend von der schriftlichen deutschen Version – interviewerbegleitet, d.h. als Face-to-Face Interview durchgeführt wird. Das würde in der praktischen Anwendung bedeuten, dass InterviewerInnen mit Kenntnis der Gebärdensprache eingesetzt werden, welche die schriftliche Fragebogenversion „übersetzen“. Diese Erhebungsform würde jedoch wieder zu einer Reihe von weiteren Problemen führen. Einige dieser Probleme sind folgende:

- Die primäre Auslegung des Instruments als Face-to-Face Interview würde den Vergleich der Ergebnisse mit schriftlichen Versionen nur mehr bedingt erlauben. Insbesondere, als im gegenständlichen Erhebungsinstrument zahlreiche persönliche Fragen (z.B. Fragen nach dem Sexualleben und der Zufriedenheit mit dem eigenen Aussehen) gestellt werden und daher mit Antwortverzerrung durch die Anwesenheit von InterviewerInnen zu rechnen ist.
- Bei der standardisierten Face-to-Face Befragung werden die Fragen und Antwortmöglichkeiten von den InterviewerInnen üblicherweise abgelesen, wodurch die weitgehende Standardisierung der Erhebungsgespräche erreicht wird. Gebärdensprache kann jedoch in der Interviewsituation nicht „abgelesen“ werden. Es wäre bestenfalls möglich, dass die InterviewerInnen die Fragen jeweils zur Interviewsituation spontan in Gebärdensprache übersetzen. Dadurch würde jedoch die Erhebungssituation durch unterschiedliche Übersetzungsvarianten eine geringe Standardisierung aufweisen.
- Die Rekrutierung der benötigten InterviewerInnen mit Kenntnissen der Gebärdensprache würde schwieriger als im herkömmlichen Erhebungsfall und insbesondere im Vergleich zur schriftlichen Version mit deutlich erhöhten Erhebungskosten verbunden sein. Da zudem die Organisation dieser Bevölkerungsgruppe – zumindest lokal – sehr

stark ausgeprägt ist, ist die Wahrscheinlichkeit der Bekanntschaft von interviewten Personen und InterviewerInnen hoch, wodurch sich die Verzerrung der Antworten und die Antworthemmung aufgrund geringerer Anonymität verstärken würde. Die Rekrutierung der InterviewerInnen aus anderen Regionen würde andererseits die Verständlichkeitsprobleme verstärken, da die Gebärdensprache – zumindest in Österreich – Dialektunterschiede aufweist.

Eine derartige Vorgehensweise schränkt darüber hinaus wiederum die Interviewsituation auf einen Kommunikationskanal – nämlich die Gebärdensprache – ein. Menschen mit Hörbehinderung sollten jedoch prinzipiell die Entscheidung über die von ihnen favorisierte Kommunikationsart selbst treffen können. Entsprechend unseren Beobachtungen ist diese Wahlfreiheit zudem situationsspezifisch zu fordern. Situationsspezifisch bedeutet, dass bei jeder Erhebungsfrage selbständig entschieden werden kann, welche Kommunikationsform gewünscht wird und es soll auch der (mehrmalige) gleichberechtigte Wechsel zwischen verschiedenen Kommunikationsformen möglich sein, insbesondere auch ein gleichzeitiges Nutzen der schriftlich und gebärdensprachlich gebotenen Information. Der Einsatz zweier Kommunikationsformen wäre zwar prinzipiell auch bei einer interviewerbegleiteten Befragung möglich, indem den Befragten gleichzeitig auch die schriftliche Form der Frage vorgelegt wird. Die beiden unterschiedlichen Kommunikationsformen würden dadurch jedoch dennoch nicht in gleichberechtigter Form zur Auswahl stehen, sondern die schriftliche Vorlage zur nachgereihten „Hilfestellung“ degradiert werden.

Statt einer Face-to-Face Form mit „Echtzeit-Übersetzung“ sollte also eine Erhebungsform entwickelt werden, welche

- die selbständige Beantwortung der Fragen weiterhin ermöglicht,
- möglichst gering von der standardisierten schriftlichen Vorgabe abweicht (sowohl inhaltlich als auch hinsichtlich der Darstellung),
- die gleichberechtigte Auswahl unterschiedlicher Kommunikationskanäle (d.h. zumindest schriftsprachlich und gebärdensprachlich) unterstützt, und
- die einfache Handhabung und den geringen Ressourceneinsatz der schriftlichen Befragung beibehält.

Diese Anforderungen konnten weitgehend durch eine interaktive computerbasierte Videobefragung erfüllt werden.

5. Computerbasierte Videobefragung

Die endgültige Realisierung des Erhebungsinstruments erfolgte als computerbasierte Befragung. Die computerbasierte Befragung (d.h. eine Befragung mittels Computer, wobei die interviewten Personen direkt am PC selbständig den Fragebogen ausfüllen) wird zwar im Vergleich zu den etablierten Erhebungsmethoden nur selten angewandt, weist jedoch insbesondere im Bereich psychologischer Tests zunehmende Verbreitung auf. Weit verbreitet hingegen ist der unterstützende Einsatz des Computers unter Verwendung von CATI-Systemen bei Face-to-Face und Telefonbefragungen. Im gegenständlichen Fall erfolgt jedoch der Einsatz des Computers ohne InterviewerInnen. Für diese Formen von Befragungen werden unterschiedliche Bezeichnungen verwendet, z.B. CASAQU (computer assisted self-administered questionnaire) oder CASI (computer assisted self interviewing). Der Computer wird in diesem Fall nicht nur zur Unterstützung oder zum Befragungsmanagement eingesetzt, sondern die Befragung selbst erfolgt mittels Computer. Der Einsatz neuer Technologien in dieser Art ist weniger weit verbreitet wenn man vom gegenwärtigen Boom der Internetbefragung absieht. Es soll an dieser Stelle jedoch kein prinzipielles Plädoyer für computerbasierte Befragungen schlechthin gehalten werden. In den meisten Anwendungsfällen standardisierter Befragungen dürfte die Paper-and-Pencil Methode eine effiziente Lösung darstellen.

Prinzipielle Gründe für eine computerbasierte Erhebung können jedoch einerseits durch ein aufwendiges Erhebungsdesign³ und andererseits durch die Notwendigkeit von multimediale Technologien gegeben sein. Multimedia-Technologien zur Datenerhebung werden beispielsweise für Medien- und Musikbefragungen eingesetzt (z.B. Müller et. al 1999) aber auch im Rahmen von psychologischen Tests bei denen ein Urteil über bestimmte soziale Situationen zu fällen ist (z.B. Chan/Schmitt 1997). Staples und Luzzo (1999) berichten beispielsweise den Einsatz von Ton- und Bildsequenzen in Tests zur Karriereplanung mittels CD-i Technologie. Die Beantwortung erfolgt jedoch auf einem Papierformular.

Der Einsatz von Multimedia-Technologie im gegenständlichen Fall erlaubt den standardisierten Einsatz der Gebärdensprache zur Befragung von Menschen mit Hörbehinderung.

Es wurde dazu das Programm Animaqu (Animated Questionnaire) entwickelt. Animaqu besitzt eine Benutzeroberfläche in der Videofenster, Textfenster, Beantwortungsfunktio-

³ Die Computerbefragung wurde beispielsweise von Gerich/Lehner (<http://www.soz.uni-linz.ac.at/aes/gerich/netzwerkanalyse/netzwerk.htm>) erfolgreich im Rahmen einer egozentrierten Netzwerkerhebung eingesetzt. Die Erhebung mit traditionellen Methoden wäre in diesem Anwendungsfall aufgrund der antwortabhängigen Fragebogensteuerung zur Befragungszeit nur schwer realisierbar gewesen.

nen und Steuerungsfunktionen enthalten sind. Im Videofenster werden Fragen und Antwortvorgaben in Gebärdensprache präsentiert. Im Textfenster wird die schriftliche Form der Frage dargestellt. Frage- und Antwortsequenzen können von den befragten Personen selbst angefordert und beliebig oft abgespielt werden. Durch die benutzergeführte Gestaltung ist damit weiterhin die Möglichkeit der selbständigen Beantwortung des Fragebogens möglich. Das Instrument ist der schriftlichen Form weitgehend äquivalent, weicht jedoch naturgemäß grundsätzlich - durch die Implementierung als Computeroberfläche - von der ursprünglichen Vorgabe der schriftlichen Befragung ab.

Die Frage der Vergleichbarkeit der Erhebungsergebnisse von Computerbefragungen und traditionellen Erhebungsmethoden kann auf der Basis bisheriger empirischer Untersuchungen nicht eindeutig beantwortet werden. Richman et al. (1999) führen zusammenfassend zahlreiche Untersuchungen an, die sie einer Metaanalyse unterzogen. Zahlreiche Befunde beziehen sich auf die Unterschiede durch das differenzierte Ausmaß sozialer Wünschbarkeit der Antworten bei traditionellen und computergestützten Erhebungsmethoden. Neben unterschiedlichen Untersuchungen, bei denen keine Methodenunterschiede festgestellt wurden, gibt es bei anderen sowohl Hinweise für Methodeneffekte im Sinne höherem oder niedrigerem Auftreten sozialer Wünschbarkeit. Begründungen für die stärkere Verzerrungsanfälligkeit bei Computerinterviews beziehen sich häufig auf das Argument des „big brother syndroms“, also dem Eindruck der befragten Personen, dass ihre Daten überprüft und in einer Datenbank gespeichert werden. Begründungen für geringere Verzerrungsanfälligkeit der computergestützten Erhebung beziehen sich zumeist darauf, dass die Atmosphäre der Privatheit und Anonymität die Antwortbereitschaft begünstigen. Richman et al. (1999) leiten jedoch aus ihren metaanalytischen Vergleichen ab, dass keine generellen Unterschiede zwischen computergestützten und traditionellen (insbesondere Paper-and-Pencil) Methoden festzustellen sind: „We found that using a computer instrument per se has no consistent effect on distortion; the effect, if any, depends on what the instrument measures and on moderating factors such as whether respondents are tested alone or in presence of others“ (Richman et al. 1999: 769). Hänsgen (in Anlehnung an Franke) erwähnt, dass Hinweise zu finden sind, die auf eine eher erhöhte Selbst- aufmerksamkeit der interviewten Personen bei computerbasierten Befragungen schließen lässt (Hänsgen 1999: 11). Er weist darauf hin, dass bei Computerbefragungen gerade in jüngeren Untersuchungen eine Tendenz zur Vereinheitlichung festzustellen ist. Eine mögliche Ursache für die zunehmend bessere Vergleichbarkeit von Paper-and-Pencil und computerbasierten Fragebögen sieht Hänsgen in der verbesserten ergonomischen Gestaltung (Farbe, Grafik-Oberfläche, verbesserte Antworteingabemöglichkeiten, etc) der Computerfragebögen

(ders.: 19f.). Für den Vergleich von Face-to-Face Interviews mit Computererhebungen ziehen Richman et al. (1999) jedoch den Schluss, dass bei letzteren tendenziell mit geringerer Verzerrung zu rechnen ist „particularly when the interviews were asking respondents to reveal highly sensitive personal behaviour, such as whether they used illegal drugs or engaged in risky sexual practices“ (dies.: 770). Auch Mead und Drasgow (1993) kommen in einer Metaanalyse zu vergleichbaren Schlussfolgerungen. Psychometrische Effekte der Computererhebung fanden sie jedoch bei empirischen Tests, für welche die Beantwortungsgeschwindigkeit von Relevanz ist. Zusammenfassend ist also festzuhalten, dass die Vergleichbarkeit zwischen Paper-and-Pencil Methode und Computererhebung von Fall zu Fall empirisch zu testen ist, jedoch die Wahrscheinlichkeit für Unterschiede dann geringer ist, wenn die Computervariante hinsichtlich der äußeren Erscheinung und der Handhabung weitgehend der Papierform entspricht.

Im gegenständlichen Fall ergeben sich Unterschiede im Vergleich zur Originalform zum einen dadurch, dass anstatt dem manuellen Ankreuzen nun die Beantwortung durch Klicken eines Buttons erfolgt. Ein weiterer Unterschied zur ursprünglichen Papierform besteht darin, dass die Befragung bildschirmorientiert erfolgt. D.h., dass jeweils nur eine Frage auf dem Bildschirm platziert wird und nicht etwa ganze Fragebatterien („One screen one item – Verfahren“). Der Grund dafür liegt im begrenzten Bildschirmplatz und der daraus folgenden Notwendigkeit des Scrollens wenn mehrere Fragen gleichzeitig präsentiert werden sollen. Dies würde zu einer aufwändigeren Bedienung und einer schlechteren Übersichtlichkeit führen.

Aufgrund dieser beiden Unterschiede ist mit divergierenden Befragungszeiten im Vergleich zur Paper-and-Pencil Methode zu rechnen: Sawaki (2001) beispielsweise führt Unterschiede in der Bearbeitungszeit vor allem auf unterschiedliche motorische Beanspruchung zurück: „'Marking a bubble' on an answer sheet in a P&P [*Paper-and-Pencil, Ann. d. A.*] test and 'pressing a button' to enter an answer on a CAT [*Computerbefragung, Ann. d. A.*] may require different motor skills. (...) keeping track of the location of the items presented as a group was part of the task in the highly-speeded P&P test, while it was not the case for the CAT version, where items were displayed one by one on a computer screen“ (Sawaki 2001: 39). Da es sich im gegenständlichen Fall jedoch nicht um Fähigkeitstests handelt, dürften diese Unterschiede hinsichtlich den psychometrischen Eigenschaften der Messung von geringer Bedeutung sein. Bei multimedialen Anwendungen kann sich darüber hinaus auch ein gegenteiliger Effekt einstellen, da die Rezeption von Ton- oder Videomedien i.d.R. langsamer erfolgt als die von Textmaterial. Johnston und Walton (1995) berichten sogar von Erhebungsproblemen durch die langsamere Befragung bei audigestützten Erhebungen: „many respondents become impatient with the slower pace of the audio as they become accustomed to pat-

terns of questioning and response alternatives“ (Johnston/Walton 1995: 317). Im Programm Animaqu kann die interviewte Person allerdings jederzeit selbst die Videosequenz stoppen, bzw. auch während der multimedialen Präsentation zur nächsten Frage wechseln. Dadurch erfolgt für literate Personen, die sich eher am schriftsprachlichen Text orientieren, theoretisch keine Verzögerung des Befragungsablaufes. Die Befragungszeiten bei Pretests mit Animaqu betragen für den WHOQOL-Bref (inkl. demografischer Fragen) durchschnittlich 19 Minuten. Personen mit guten Lesefähigkeiten brauchten für die Beantwortung etwa fünf Minuten, Personen, welche keine oder geringe literate Fähigkeiten besitzen bis zu 45 Minuten. D.h. aus den bisherigen Erfahrungen kann davon ausgegangen werden, dass die Befragungszeiten bei lesegewandten Personen mit der Paper-and-Pencil Version vergleichbar⁴ sind.

Der letztlich wesentliche Unterschied liegt jedoch natürlich in der Einbindung der Videosequenzen in Gebärdensprache. Aber gerade in diesem wesentlichen Unterschied zur schriftlichen Version liegt der Hauptgrund, warum wir meinen, dass die computerbasierte Erhebungsform der standardisierten Originalversion eher entspricht als eine Face-to-Face Befragung mit „Echtzeit-Übersetzung“ in Gebärdensprache: Es ist möglich durch den Einsatz von Expertengruppen (gehörlose Menschen, Native-Speaker, Dolmetscher etc.) aus einer Vielzahl möglicher Übersetzungen die geeignetsten auszuwählen. Im konkreten Fall wurde zusätzlich zur Expertenselektion und Pretests mit Beobachtung eine mehrstufige Übersetzungs- und Rückübersetzungsprozedur angewandt, um die einzelnen Gebärdendarstellungen hinsichtlich ihrer Allgemeinverständlichkeit und sinngemäßen Richtigkeit zu prüfen. D.h., es wurden verschiedene Videoaufnahmen verschiedener Personen aufgezeichnet. Die Videos wurden anschließend von Experten nach dem präzisen Einsatz von Gebärden selektiert. Zudem wurden die Gebärden in Textsprache übersetzt und anschließend wieder rückübersetzt. Auf der Basis der Übereinstimmung von Übersetzung und Rückübersetzung wurden geeignete Videosequenzen selektiert.

Die Frage, die sich aufgrund der Videoeinbindung im besonderen stellt, nämlich inwiefern Multimedia-Befragungen mit Paper-and-Pencil Versionen vergleichbar sind ist bislang natürlich kaum behandelt worden. Staples und Luzzo (1999) beispielsweise vergleichen Paper-and-Pencil Versionen von Career Assessment Inventaren mit Multimedia-Versionen derselben Instrumente. Die Ergebnisse der Vergleiche zeigen die prinzipielle Äquivalenz der Testergebnisse. Einschränkend ist jedoch festzuhalten, dass in diesem Fall keine Videosequenzen sondern lediglich Bilderfolgen und Tonsequenzen mittels CD-i präsentiert wurden und die Beantwortung mittels Papierformular erfolgte.

⁴ Die Bearbeitungsdauer beim deutschsprachigen WHOQOL-Bref beträgt zwischen 5 und 10 Minuten (Angermeyer et al. 2000, 25).

Im Vergleich zu einer Face-to-Face Variante kann bei Animaqu wie bei der schriftlichen Befragung auf InterviewerInnen verzichtet werden. Wie bei der schriftlichen Selbstbeantwortung kann mit Animaqu der Frageinhalt und die Antwortmöglichkeit beliebig oft rezipiert werden. D.h. so wie der Text beliebig oft gelesen werden kann, können auch die Videosequenzen beliebig oft gestartet und angehalten werden. Durch die simultane Präsentation von Text und Bild werden Gebärdensprache und Schriftsprache gleichberechtigt eingesetzt. Es wird lediglich ein Versuchsleiter benötigt, um das Untersuchungsziel und die Bedienung zu erläutern. Durch den Entfall der InterviewerInnen werden auch die Interviewereffekte und das Anonymitätsproblem weitestgehend behoben. Die Darsteller, welche die Fragen in den Videos präsentieren sind selbst gehörlose Menschen. Die Frage, ob und in welcher Weise sich dadurch möglicherweise „indirekte“ Interviewereffekte ergeben, kann gegenwärtig nicht geklärt werden. Weisband und Kiesler (1996) kommen jedoch im Rahmen einer Metaanalyse zu dem Schluss, dass die Antwortbekennung in Computerbefragungen höher als in Face-to-Face und auch Paper-and-Pencil Situationen ist. Auf der anderen Seite stellen sie jedoch fest, dass durch die Simulation sozialer Situationen in Computerfragebögen die Antwortbekennung wieder abnimmt und konstatieren zusammenfassend: „Computer interfaces that partially mimic social situations may offer a distorted version of these situations, and change people's responses unexpectedly“. Obwohl bei der Befragung mit Animaqu keine sozialen Situationen im eigentlichen Sinne „nachgestellt“ werden, sind also dennoch Verzerrungen durch die Videoeinbindung nicht auszuschließen. Da jedoch sämtlichen Befragten dieselben Videos präsentiert werden, würde es sich weitgehend um konstante Effekte handeln.

Richman et al. (1999) plädieren wie bereits erwähnt dafür, dass Computererhebungen möglichst so gestaltet werden sollten, dass sie in ihrer äußeren Erscheinungsform schriftlichen Befragungen weitgehend entsprechen: „the more a computer instrument resembles a traditional instrument, the more the two instruments should produce similar responses“ (Richman et al. 1999: 756). Für den Anwendungsfall der Videobefragung ist dieses Ziel jedoch nur schwer zu verfolgen.

Hinsichtlich der Handhabung des Instruments ist zunächst prinzipiell festzuhalten, dass die Beantwortung der Computerversion des Fragebogens nicht schwieriger ist als die der Papierform. Dies setzt jedoch voraus, dass die Gestaltung der Benutzeroberfläche möglichst intuitiv erstellt wird. Insbesondere, da durch die Videosteuerung und die Funktionen zum Navigieren des Fragebogens zusätzliche „Bedienungselemente“ im Vergleich zur schriftlichen Befragung erforderlich sind. Die Vergleichbarkeit des äußeren Erscheinungsbildes von Computerbefragung und Papierversion kann jedoch im gegenständlichen Fall durch die Verwendung der Gebärdensprache grundsätzlich nicht erreicht werden. Die Gestaltung der Oberfläche wurde nicht zuletzt deshalb in einem intensiven Austauschpro-

zess mit gehörlosen Menschen vorgenommen. Ziel dieses Abstimmungsprozesses war es jedoch weniger, das Erscheinungsbild der Computerbefragung an das der schriftlichen Form anzupassen, als vielmehr die Handhabung des Instruments bestmöglich zu gestalten. Mit wesentlichen Handhabungsproblemen ist jedoch bei jenen Personen zu rechnen, die grundsätzlich reservierte Haltungen gegenüber Computern haben. Dies stellt aus unserer Sicht auch das Hauptproblem bei der Vergleichbarkeit der Erhebungsmethoden dar und bedarf weiterer entsprechender Untersuchungen. Die prinzipielle Hemmschwelle gegenüber dem Medium Computer kann jedoch möglicherweise durch den Einsatz entsprechender Eingabemedien entschärft werden (vergleiche dazu die nachfolgenden Ausführungen).

6. Vergleichender Ressourceneinsatz

Booth-Kewley et al. (1992) ziehen aus den Ergebnissen Ihrer Methodenvergleiche, dass computer- und Paper-and-Pencil Methoden vergleichbare Ergebnisse liefern, den Schluss, dass computergestützte Methoden - mit Vorbehalten - bevorzugt eingesetzt werden können: „This finding indicates that, where financial and logistical considerations allow, researchers are justified in using computers instead of paper-and-pencil to gather attitude data because of the many previously cited advantages of computerized data collection“ (Booth-Keweley et al. 1992: 565).

Der Ressourceneinsatz der computerbasierten Erhebung ist differenziert zu betrachten: Zunächst ist natürlich die Entwicklung des Erhebungsinstrumentes um ein vielfaches aufwändiger als die einer schriftlichen Befragung. Hinzu kommt eine im Vergleich zur schriftlichen Form ebenfalls verlängerte Testphase. Diese Schritte könnten in Zukunft jedoch deutlich reduziert werden, wenn für diese Zwecke Standardsoftware entwickelt wird.⁵ In der Feldphase selbst besteht der Ressourcenaufwand in der Verfügbarkeit der Geräte. Je nach Befragungsmöglichkeiten und Modalitäten müssen ein oder mehrere Desktop-Rechner oder Notebooks zur Verfügung stehen. Im gegenständlichen Fall besteht die Möglichkeit, die Erhebung in den Vereinslokalitäten der örtlichen Gehörlosenvereine durchzuführen, wodurch der Aufwand aufgrund der bestehenden Infrastruktur als gering zu bewerten ist. Im Vergleich zur Face-to-Face Erhebung entfallen die Interviewerkosten, im Vergleich zur schriftlichen Befragung entfallen die Vervielfältigungskosten der Fragebögen. Hinsichtlich der Datenerfassung bietet die computerbasierte Erhebung natür-

⁵ Animaqu wurde als Prototyp auf MS-Access in Visual Basic programmiert und ist auf den hier eingesetzten Fragebogen beschränkt. Eine Verallgemeinerung für die Anwendung auf nahezu beliebige Befragungen wäre jedoch denkbar.

lich immer Ressourceneinsparung, da die Daten zeitgleich mit der Erhebung elektronisch erfasst werden. Neben der Kosteneinsparung ergibt sich daraus zusätzlich der Vorteil einer Fehlerreduktion bei der Datenerfassung wie er auch bei CAPI und CATI - Anwendungen bekannt ist. Es ist also aus den jeweiligen Erhebungsbedingungen und der zukünftigen Softwareentwicklung abhängig, wie der jeweilige Ressourceneinsatz zu bewerten ist.

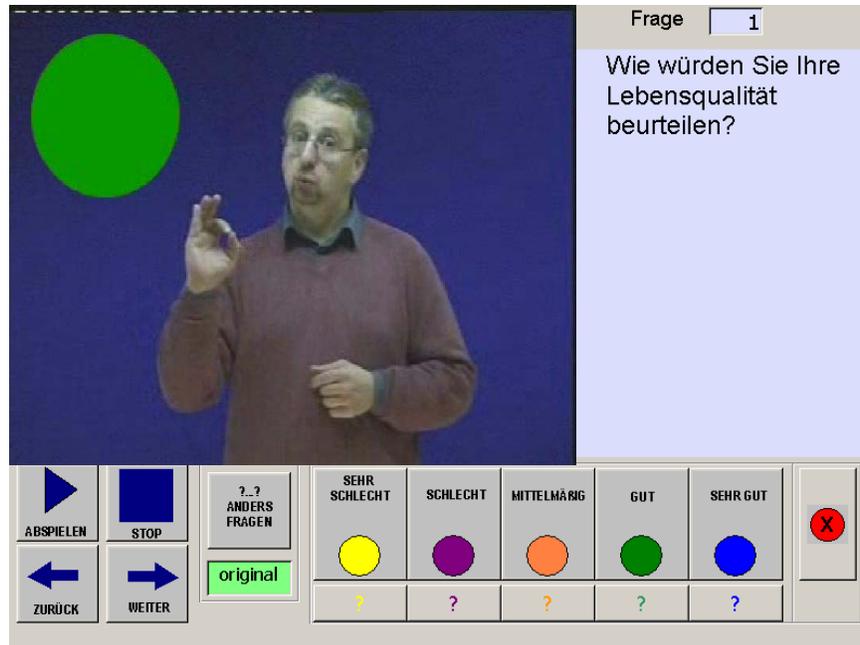
7. Die konkrete Umsetzung

Die zentrale Funktionalität ist in Abbildung 1 dargestellt. Eine besondere Schwierigkeit bei der Instrumententwicklung ergab sich zum einen dadurch, dass eine möglichst übersichtliche und – insbesondere für Personen, die um Umgang mit Computern wenig bzw. keine Erfahrungen haben – einfache Bedienung bei gleichzeitig beträchtlicher Funktionsmächtigkeit der Umsetzung erreicht werden sollte. Des weiteren ist eine bestimmte Mindestgröße für die Videodarstellung zu reservieren, wodurch Platzprobleme am Bildschirm resultieren. Neben der Anforderung an die Video-Mindestgröße musste sowohl eine gut lesbare Schriftgröße gewählt, als auch relativ große Antwortbuttons zur einfachen und klaren Bedienbarkeit des computerbasierten Fragebogens gewählt werden. Diese Bedingungen an das Erhebungsinstrument hatten zur Folge, dass neben den - bereits oben erwähnten - Navigationselementen, dem Video selbst und der eigentlichen schriftsprachlichen Frage maximal fünf Antwortmöglichkeiten (fünf Antwortbuttons) je Frage zur Verfügung gestellt werden können. Die gewählten Erhebungsinstrumente zur Lebensqualität erfüllen dabei diese Anforderungen. Für einzelne demographische Fragen stellt die geringe Anzahl der möglichen Ausprägungen jedoch eine Einschränkung dar.

Gerade bei der gewählten Videogröße bzw. dem Videoformat war ein Kompromiss zwischen der Qualität der Videodarstellung und dem Kompressionsgrad - d.h. der Dateigröße - der einzelnen Videofiles zu finden. Die Videogröße und -qualität muss so gewählt werden, dass die Gebärden klar und deutlich erkennbar sind, damit soweit wie möglich Fehlinterpretationen aufgrund des gewählten Videoformats ausgeschlossen werden. Sämtliche Details (Fingerbewegung, Lippenbewegung, Mimik, etc.) müssen klar erkennbar sein, damit die Gebärdendarstellung eindeutig rezipierbar ist. Die Dateigrößen der jeweiligen Videofiles dürfen im Gegenzug ein gewisses Maß nicht übersteigen, da ansonsten die bei der Befragung eingesetzten Computer überfordert werden. In diesem

Fall werden die Videos zum Teil ruckartig dargestellt und der Übergang von einer zur nächsten Frage aufgrund des Ladevorganges stark verzögert.⁶

Abbildung 1: Zentrale Funktionalität des Videofragebogens



Dargestellt ist der Bildschirm für die erste Frage des WHOQOL-BREF. Die Präsentation der übrigen Kernfragen erfolgt in gleicher Weise. Bei einzelnen in dieser Untersuchung eingesetzten Zusatzfragen wurde die Gestaltung an die Erfordernisse angepasst.⁷ Wesentlich ist die Konstanz der Bildschirmgestaltung für alle Fragen. Dies wurde durch eine farblich

⁶ Folgende Kennzahlen haben sich bewährt: Videogröße (Frame size): 384x288 mit 25.000fps. Durchschnittliche Bitrate: 2365kilobits/sec. Als Videoformat wurde das avi-Format und als Kompressionsverfahren das „mpeg4 v1-layer“ gewählt. Sehr gute Ergebnisse bei der Darstellung sind bereits mit älteren Computern möglich (z.B. Pentium 166 mit 15"-Monitor). Notwendige Festplattenkapazität: rund 135MB (für die Fragen bzw. Antworten der drei eingesetzten Erhebungsinstrumente und den demographischen Fragen).

⁷ Dies betrifft Fragen mit weniger Antwortmöglichkeiten und eine grundsätzlich andere Gestaltung bei der Frage nach dem Geburtsjahr.

che Zuordnung der Antwortkategorien unterstrichen. Die zugeordneten Farben der Antwortkategorien werden auch in der Videosequenz der entsprechenden Antwortmöglichkeit eingeblendet. Die Textierung der Antwortkategorien wird jeweils an die Frageformulierung angepasst, wobei jedoch die farblichen Symbole darunter gleich bleiben.

Beim Betreten einer Frageseite wird die Videosequenz der Frage automatisch gestartet. Jedes Fragevideo beinhaltet den gesamten Frageinhalt inklusive aller Antwortkategorien. Die einzelnen Antwortkategorien können jedoch auch einzeln über die jeweiligen Buttons mit den Fragezeichen, die sich unterhalb der Beantwortungsbuttons befinden abgespielt werden. Diese Lösung zur Präsentation der Antwortkategorien hat sich durch die praktische Erprobung als zweckmäßig erwiesen. Andere Lösungen wie z.B. die Präsentation der Antwortvideos in einem weiteren Fenster führen zu geringerer Übersichtlichkeit bzw. zu Platzproblemen in der Darstellung.

Im Zusammenhang mit den Antwortvorgaben ist auf eine weitere Besonderheit hinzuweisen, die bei der Anwendung von Gebärdensprache in standardisierten Befragungen auftritt: Bei standardisierten schriftlichen Befragungen werden zumeist Fragebatterien eingesetzt. Eine wesentliche Eigenschaft von Fragebatterien liegt in einer gewissen „Rationalisierung“ der Frageformulierung und auch der Beantwortung, da zahlreiche Frage-Items zusammengefasst mit wortidentischen Antwortvorgaben verbunden werden. Die Darstellung erfolgt zumeist in Listenform, wobei die Antwortvorgaben in der Regel nur einmal am Listenkopf präsentiert werden. So besteht z.B. die schriftliche Version des WHOQOL-BREF aus sieben Fragebatterien, wobei insgesamt lediglich fünf unterschiedliche Formen von Antwortskalen verwendet werden. Die gestraffte Art der Präsentation von Fragen in Form einer Fragebatterie ist mit der computerbasierten Videobefragung für Menschen mit Hörbehinderung aus zwei Gründen nicht realisierbar. Der erste Grund liegt in der bildschirmorientierten Befragungsgestaltung. Da das Instrument so konzipiert ist, dass immer nur eine Frage am Bildschirm dargestellt wird, besteht keine sinnvolle Möglichkeit Fragebatterien gemeinsam mit zugehörigen Fragevideos zusammen auf einer Bildschirmseite zu platzieren. Der zweite Grund liegt im Unterschied zwischen Schrift und Gebärdensprache: Ein wesentlicher Teil der Semantik von Gebärdensprache wird über die Ebene der Mimik und des emotionalen Körperausdrucks vermittelt. Daraus folgt, dass Antwortvorgabe und Frageinhalt nur bedingt voneinander isolierbar sind, was jedoch eine Voraussetzung für den Anwendungsfall der Fragebatterie darstellt, wenn dieselben Antwortkategorien für mehrere Fragen gleichzeitig verwendet werden. Die Antwortvorgaben sind für jedes Item einer Fragebatterie wortident, jedoch nicht notwendigerweise sinnident. Dies soll anhand zweier Items einer Fragebatterie des WHOQOL-BREF verdeutlicht werden:

Frage 3 und Frage 5 sind in einer Fragebatterie zusammengefasst und haben in der deutschen Version folgenden Wortlaut:

		überhaupt nicht	ein wenig	mittelmäßig	ziemlich	äußerst
3	Wie stark werden Sie durch Schmerzen daran gehindert, notwendige Dinge zu tun?	1	2	3	4	5
5	Wie gut können Sie Ihr Leben genießen?	1	2	3	4	5

Insbesondere da die beiden Items gegensinnig gepolt sind, können bei der Übersetzung in Gebärdensprache nicht dieselben Antwortvorgaben für beide Items verwendet werden, da sie unterschiedlich dargestellt werden. Die Vermittlung der Antwortkategorie „äußerst“ wird im Fall des Items Nummer 5 mit strahlendem Gesicht und emotional freudigem Körperausdruck dargestellt, im Fall von Item 3 mit Leidensausdruck. Es ist daher in der Regel nicht möglich, dieselben Antwortkategorien in Gebärdensprache für unterschiedliche Frageformulierungen einzusetzen. Aus diesem Grund reichte es also nicht aus, lediglich die fünf verschiedenen Antwortskalen des WHOQOL-BREF in Gebärdensprache zu übersetzen, sondern es wurden für jede Frage getrennt die Antwortkategorien aufgezeichnet.

Die Beantwortung der Fragen erfolgt durch die Antwortbuttons, welche bei Betätigung zur Signalisierung der gewählten Antwort gedrückt bleiben. Die Auswahl der gewünschten Antwortkategorie kann dabei jederzeit, unabhängig davon, ob das Fragevideo bereits vollständig abgespielt wurde, gewählt werden.

Die Videosteuerung erfolgt durch die beiden Funktionen „abspielen“ und „stop“. Die Navigation durch den Fragebogen, d.h. das „blättern“ durch die einzelnen Fragen des Fragebogens kann durch die Funktionen „zurück“ und „weiter“ vorgenommen werden. Die Möglichkeit, zu bereits beantworteten Fragen zurückzuspringen und auch bei Bedarf Korrekturen an der Beantwortung vorzunehmen wurde explizit berücksichtigt, da dies auch bei der schriftlichen Befragung möglich ist. Auch in der Literatur wird überwiegend die Möglichkeit des Navigierens im Fragebogen (backtrack) und Überspringen von Items ohne sie zu beantworten (skip items) im Vergleich zu starren Designs (lock-step) bevorzugt: „Researchers have argued that if respondents are constrained by the computer in the ways they can look at items and answers (e.g., by not being able to skip items, not having the option of answering 'don't know', or not being allowed to backtrack), they may feel self-conscious or wary“. (Richman et al. 1999, 756). Die Autoren fanden Hinweise dafür, dass bei backtrack Methoden die Wahrscheinlichkeit höher ist, dass soziale Wünschbarkeit bei der Beantwortung reduziert wird. Booth-Kewelely et al. (1992) kommen bei

Methodenvergleichen zwar zu dem Schluss, dass keine Antwortunterschiede zwischen backtrack und lock-step Verfahren bestehen, geben jedoch dennoch der backtrack Methoden den Vorzug, da 61% der Befragten von dieser Möglichkeit zumindest einmal und 45% mehrmals Gebrauch machten.

In einer früheren Version bestand auch die Möglichkeit ohne Beantwortung der aktuellen Frage durch den Fragebogen zu navigieren, da auch dies der schriftlichen Befragungssituation entspricht. In den Praxistests hat sich jedoch gezeigt, dass in manchen Fällen irrtümlicherweise die Fragezeichenbuttons (zur Präsentation der Antwortvideos) anstatt der Antwortbuttons gedrückt werden. Die Folge war eine irrtümlicherweise nicht beantwortete Frage. Dieses Problem wurde dadurch gelöst, indem - abweichend vom Originalfragebogen - die Antwortenthaltung als eigene Antwortkategorie eingeführt wurde („x“-Button am rechten unteren Bildschirmrand). Zusätzlich ist nun die Navigation im Fragebogen nur dann möglich, wenn die aktuelle Frage beantwortet (oder die explizite Antwortenthaltung gewählt) wurde, um versehentliche Nichtbeantwortung zu verhindern.

Eine weitere Funktion betrifft den Button „anders fragen“. Diese Funktion wurde von Gehörlosen und den Projektpartnern der Gehörlosenambulanz angeregt. Mit dieser kann die schriftsprachliche Fragetextierung in einer alternativen - d.h. kürzeren und vereinfachten - Formulierung angezeigt werden. Zur Kennzeichnung, welche Frageversion gerade angezeigt wird, erscheint in einem Feld unter diesem Button der Hinweis „original“ oder „kurz“. Nach erfolgter Beantwortung der Frage und darauffolgendem Wechsel zur nächsten Frage wird jeweils wieder der Originalfragetext angezeigt, wobei anschließend wiederum auf die Kurzfassung umgeschaltet werden kann.

Aus dieser Zusatzfunktionalität entsteht natürlich eine weitere Abweichung vom schriftlichen Originalfragebogen. Es wird daher auch während der Befragung protokolliert, wie oft eine befragte Person zwischen den unterschiedlichen Fragetexten wechselt.

8. Die Erhebungssituation

Wie bereits beschrieben, ist bei der Erhebung mit Animaqu kein Interviewereinsatz nötig. Im Vergleich zu einer schriftlichen Befragung ist jedoch eine längere Vorbereitungsphase für die befragten Personen nötig. Daraus folgt, dass den Versuchsleitern im Vergleich zur schriftlichen Befragung eine bedeutendere Rolle zukommt. Für die spätere Befragungssituation ist entscheidend auf welche Art und Weise die befragten Personen durch den Versuchsleiter mit dem Instrument vertraut gemacht werden. Neben der grundsätzlichen Erläuterung der Bedienung besteht eine wichtige Funktion der Vorbereitungsphase darin, bestehende Berührungängste mit dem Computer zu überwinden. In Pretests

hat sich bestätigt, dass die Einschulung durch manche Versuchsleiter mangelhaft war, mit der Folge, dass den Befragten beispielsweise wichtige Bedienungsfunktionen nicht vermittelt wurden oder die Erklärungen unstrukturiert erfolgten, sodass die befragten Personen verunsichert wurden. Um die Befragungssituation auch hinsichtlich der Vorbereitungsphase möglichst standardisiert zu gestalten, wurde ein Manual für Versuchsleiter erstellt. Das Manual umfasst eine vorgeschriebene Abfolge, nach der die befragten Personen mit der Bedienung vertraut gemacht werden. Die Erklärung der Bedienung selbst wird durch eine „Trainingsversion“ von Animaqu unterstützt. Die Trainingsversion kann beliebig lange getestet und auch jederzeit beendet werden und besteht aus Trainingsfragen, die natürlich nicht zur eigentlichen Hauptbefragung gehören. Sie ist mit der Hauptbefragung im Erscheinungsbild und den darin vorkommenden Bedien- bzw. Antwortmöglichkeiten identisch.

Versuchsleiter und befragte Personen entscheiden über den Zeitpunkt, wann die Hauptbefragung begonnen wird. Falls durch eine Fehlbedienung irrtümlicherweise die Hauptbefragung gestartet wurde, oder die Beantwortung der Hauptbefragung aus irgendwelchen Gründen ungültig ist, kann im Anschluss die Status Information abgerufen werden. Aus der Status Information kann die ID-Nummer des letzten Datensatzes ermittelt und in ein Fehlerprotokoll aufgenommen werden.

9. Die Auswahl der Beantwortungsformen

Zur Bedienung können unterschiedliche Eingabeformen eingesetzt werden. Da die meisten Personen bereits computererfahren sind, kann zumeist die Maus als Eingabemedium verwendet werden. Die Computermaus hat – wie der weiter unten beschriebene Touchscreen – zudem den Vorteil, dass dabei „die visuelle Orientierung ständig auf den Bildschirm fokussiert bleibt“ (Hänsgen 1999: 22). Zusätzlich sind alle Funktionen auch über einzelne Tasten der Tastatur ansteuerbar. Dazu werden die jeweils belegten Tasten etikettiert. Die Etiketten sind mit den Symbolen der Bildschirm-Buttons bzw. den Farbkreisen der Antworttasten bedruckt. Die Verwendung von Spezialtastaturen mit einer reduzierten Anzahl von Tasten wäre zudem als Eingabemedium denkbar. Als weitere Bedienungsmöglichkeit wird die Touchscreen-Technologie genutzt. Die Bedienung mittels Touchscreen stellt natürlich die einfachste und intuitivste Lösung dar (vgl. z.B. Weichbold 2002; Swoboda 1996), ist jedoch mit einem höheren Ressourcenaufwand in der Feldphase verbunden, da Touchscreen-Monitore teurer als herkömmliche Bildschirme sind. Touchscreens zeichnen sich dadurch aus, dass die Ein- und die Ausgabe der Informationen eine einzelne kompakte Einheit bildet, die zudem die Augen-Hand-Koordination erleichtert. Das Aufrechterhalten eines ständigen Blickkontaktes ist

insbesondere für Menschen, die üblicherweise in Gebärdensprache kommunizieren von besonderer Bedeutung. Im Rahmen von Pretests erwies sich auch gerade bei wenig computererfahrenen Personen der Blickwechsel von Bildschirm zu Tastatur z.T. als problematisch⁸, weshalb sich die Bedienung mittels Touchscreen als die zuverlässigste Lösung herausgestellt hat. Diese Befunde sind auch ein Beleg für die von Hänsgen (1999) angeführte höhere Vergleichbarkeit von Papier- und Computerbefragungen aufgrund besserer ergonomischer Gestaltung der elektronischen Erhebungen. Eingabemedien wie vor allem Touchscreens, die der intuitiven Bedienung entgegen kommen, können die Güte der Erhebungen verbessern.

10. Zusammenfassende Beurteilung des Erhebungsinstrumentes

Durch die computerbasierte Videobefragung ist es gelungen, eine - der schriftlichen standardisierten Befragung weitgehend äquivalente - Befragungsmethode zu entwickeln, die eine Erweiterung der möglichen Zielgruppe bei standardisierten Erhebungen erlaubt. Die Anwendung ist theoretisch nicht auf die Gruppe der Menschen mit Hörbehinderungen beschränkt.

Generell gilt für die empirische Sozialforschung, dass Erhebungsinstrumente eine „box of tools“ darstellen, jedoch jene Methoden gewählt werden sollten, die für den Anwendungsfall am geeignetsten sind. So sollte unserer Ansicht nach auch die Computerbefragung und insbesondere die Multimedia-Befragung nicht willkürlich eingesetzt werden. Obwohl bisherige Vergleichsuntersuchungen keine einheitlichen Schlussfolgerungen zulassen, besteht dennoch die Wahrscheinlichkeit zahlreicher Verzerrungsanfälligkeiten. Denkbar wären beispielsweise jedoch Implementierungen für verschiedene Personengruppen, die über eingeschränkte schriftsprachliche Kommunikationsfähigkeiten verfügen (z.B. Kinder, Personen mit eingeschränkten Lese- oder Schreibkenntnissen etc.). Computergestützte Kommunikationstechnologien für Menschen mit Sehbehinderungen sind bereits weit verbreitet. Durch die Verwendung von Tondokumenten würde auch die standardisierte Befragung nach dem dargestellten Muster ermöglicht. Darüber hinaus bieten computergestützte Verfahren durch die Realisierung eines dynamischen Ablaufes zur Befragungszeit die Möglichkeit komplexe Erhebungen zu realisieren. Obwohl unter bestimmten Voraussetzungen auch die Anwendung von Internetbefragungen sinnvoll

8 In der Folge wurden beispielsweise irrtümlicherweise die falschen Antwortkategorien gewählt. Weiters trat durch den verzögerten Blickwechsel von Tastatur zu Bildschirm das Problem auf, dass beim Start der Videos mittels Tastatur der Beginn der Videosequenz am Bildschirm übersehen wurde.

scheint, wird der Einsatz neuer Medien in der Datenerhebung unserer Ansicht nach zu sehr auf diese Einsatzebene eingeschränkt. Es wird dabei zuweilen außer acht gelassen, dass die medialen Vorteile im Rahmen von Offline-Erhebungen nicht nur ebenfalls ausgenutzt werden können, sondern, dass darüber hinaus die zahlreichen Probleme der Online-Erhebungen (Stichprobenproblematik, Situationskontrolle, Übertragungsgeschwindigkeit etc.) vermieden werden. Es konnte anhand der hier präsentierten Methode demonstriert werden, dass der Einsatz computergestützter Erhebungen eine sinnvolle Ergänzung, nicht jedoch den Ersatz des traditionellen Erhebungsrepertoire darstellt.

Korrespondenzadresse

*Dr. Joachim Gerich
Institut für Soziologie
Abteilung für Empirische Sozialforschung
Johannes Kepler Universität Linz
Altenberger Straße 69
A-4040 Linz
email: joachim.gerich@jku.at*

Literatur

- Angermeyer, M.C./Kilian, R./Matschinger, H., 2000: WHOQOL – 100 und WHOQOL – BREF. Handbuch für die deutschsprachige Version der WHO Instrumente zur Erfassung der Lebensqualität. Göttingen.
- Atteslander, P., 1993: Methoden der empirischen Sozialforschung. 7. Auflage, Berlin u.a.
- Booth-Kewley, S./Edwards, J.E./Rosenfeld, P., 1992: Impression Management, Social Desirability, and Computer Administration of Attitude Questionnaires: Does the Computer Make a Difference? *Journal of Applied Psychology*, vol. 77, no. 4: 562-566.
- Chan, D./Schmitt, N., 1997: Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, vol. 82: 143-159.
- Conrad, R. 1979: *The deaf school child*. London.
- Derogatis, L.R./Spencer, P.M. 1982: *Brief Symptom Inventory: Administration, scoring and procederes manual I*. Baltimore.
- Dotter, F., 1996: Computer for the deaf (and hearing impaired): Towards an integrated solution from a linguistic standpoint. S. 205-210 in: Klaus, J. et al. (Eds.), *Interdisciplinary aspects on computers helping people with special needs*. Linz.
- Gelter, I., 1989: Wortschatz und Lesefähigkeit gehörloser Schüler. *Der Sprachheilpädagoge*, H.3: 37-42.

- Goldberg, D.P., 1978: Manual of the General Health Questionnaire. Windsor.
- Hänsgen, K.-D., 1999: Computereinsatz in der Psychodiagnostik – Stand und mögliche Perspektiven. Forschungsbericht Nr. 141. Freiburg.
- Johnston, J./Walton, C., 1995: Reducing Response Effects for Sensitive Questions: A Computer Assisted Self Interview with Audio. *Social Science Computer Review* 13:3: 304-319.
- Kannapell, B., 1989: Inside the Deaf Community. In: Wilcox, Sherman (Ed.): *American Deaf Community*. Burtonsville.
- Lucas, C. (Ed.) 1996: *Multicultural Aspects of Sociolinguistics in Deaf Communities*. Washington D.C.
- Mead, A.D./Drasgow, F., 1993: Equivalence of Computerized and Paper-and-Pencil Cognitive Ability Tests: A Meta-Analysis. *Psychological Bulletin*, vol. 114, no. 3: 449-458.
- Müller, R./Dongus, N./Ebert, S./Glogner, P./Kreutle, A., 1999: Identitätskonstruktion mit Medien und Musik. Ein empirisches Forschungsprogramm mit MultiMedia-Fragebögen. *Medien Praktisch*, Jg. 22, H.1: 26-30.
- Richman, W.L./Weisband, S./Kiesler, S./Drasgow, F., 1999: A Meta-Analytic Study of Social Desirability Distortion in Computer-Administered Questionnaires, Traditional Questionnaires, and Interviews. *Journal of Applied Psychology*, vol. 84, no. 5, pp. 754-775.
- Sawaki, Y., 2001: Comparability of conventional and computerized tests of reading in a second language. *Language Learning & Technology*, vol. 5, Nr.2, pp. 38-59.
- Schein, J.D., 1979: Society and culture of hearing impaired people. In: Bradford L.J./Hardy W.G. (eds.), *Hearing and Hearing Impairment*. New York.
- Staples, J.G./ Luzzo, D.A., 1999: Measurement Comparability of Paper-and-Pencil and Multimedia Vocational Assessments. ACT Research Report Series 99-1. Iowa City.
- Swoboda, B., 1996: *Interaktive Medien am Point of Sale. Verhaltenswissenschaftliche Analyse der Wirkung multimedialer Systeme*. Wiesbaden. (Univ. Diss.)
- The WHOQOL Group 1998: Development of the World Health Organization WHOQOL-BREF quality of life assessment. *Psychological Medicine*, 28: 551-558.
- Weichbold, M., 2002: Touchscreenbefragungen. Erste Erfahrungen zu einer neuen Form der Datenerhebung. *Österreichische Zeitschrift für Soziologie* 3/2002: 93-107.
- Weisband, S./Kiesler, S., 1996: Self Disclosure on Computer Forms: Meta-Analysis and Implications. Paper presented at the conference on Human Factors in Computing Systems, April 1996, Vancouver.
http://www.acm.org/sigchi/chi96/proceedings/papers/Weisband/sw_txt.htm (1.7.2002)
- Wilcox, S., (Ed.) 1989: *American Deaf Community*. Burtonsville.

UMFRAGEEINSTELLUNG UND UMFRAGEERFAHRUNG¹

Die relative Bedeutung unterschiedlicher Aspekte der Interviewerfahrung für die generalisierte Umfrageeinstellung

VOLKER STOCKÉ & BETTINA LANGFELDT

Der Artikel geht der Frage nach, ob sich die Bewertung vergangener Interviewerfahrungen auf die generalisierten Einstellungen gegenüber Umfragen auswirkt. Dabei steht insbesondere die Frage nach der relativen Erklärungskraft verschiedener Bewertungsdimensionen dieser Erfahrung im Mittelpunkt. Nach den Ergebnissen einer Untersuchung mit einer lokalen Zufallsstichprobe von Befragten lassen sich im ersten Schritt drei orthogonale Dimensionen dieser Bewertungen identifizieren. Demnach unterscheiden die Befragten das Ausmaß der aufgetretenen Belastung durch das Interview, den „Unterhaltungswert“ der Befragung sowie die Irritation durch verwirrende Frageformulierungen. Als zweites Ergebnis der Analyse hat sich der Erhebungsmodus bei der letzten Interviewteilnahme als bedeutsamer Prädiktor der aktuellen Bewertung von Umfragen erwiesen. So bewirkt ein face-to-face Interview bei den Befragten zuhause eine signifikant negativere Umfrageeinstellung als weniger die Privatsphäre beeinträchtigende Befragungsarten. Als drittes und zentrales Ergebnis der Untersuchung lässt sich festhalten, dass lediglich das Ausmaß der bei der letzten Interviewteilnahme erlebten Belastungen einen signifikanten Einfluss auf die allgemeine Umfragebewertung ausübt. Dieser Zusammenhang wird zusätzlich durch die mittels Antwortlatenzen operationalisierte kognitive Verfügbarkeit der Bewertungsurteile qualifiziert. Die subjektive Bewertung der in der Vergangenheit erlebten Belastung durch Interviews wirkt sich dann stärker auf die aktuelle Umfrageeinstellung aus, wenn diese Bewertungen zunehmend schneller geäußert werden. Demnach haben kognitiv zugängliche und saliente

¹ Wir danken zwei anonymen Gutachtern dieser Zeitschrift und ganz besonders Hartmut Esser für hilfreiche Kommentare und Anregungen. Birgit Becker, Christian Hunkler und Tobias Stark waren eine große Unterstützung bei der Datenaufbereitung und Manuskripterstellung. Der Beitrag wurde durch finanzielle Förderung der Deutschen Forschungsgemeinschaft an den Sonderforschungsbereich 504 der Universität Mannheim ermöglicht.

Erfahrungen mit belastenden Interviews besonders starken Einfluss auf die generalisierte Umfrageeinstellung und damit auf die mit dieser verbundenen Teilnahmebereitschaft an Umfragen in der Zukunft.

In the following article we analyze whether and to which degree respondents' evaluations of past interview experiences affect their generalized attitudes towards surveys. In particular, our study compares the relative significance of different evaluation dimensions. Our *first* result from a local survey based on a random probability sample indicates that the respondents judge their past survey experiences on orthogonal evaluation dimensions: the burden caused by interview participation, the "entertainment value" of an interview, and the irritation due to confusing question wordings. As a *second* result, the mode of administration of the last interview proved to be a significant predictor of the evaluation of surveys in general. A face-to-face interview at the respondents' home is found to have a more negative effect on survey attitudes compared with modes which invade privacy to a lesser degree. The *third* and most important finding of the analysis shows that the amount of burden experienced during the last interview has a significant effect on the generalized evaluation of surveys. Whether the last survey interview is judged positive or negative on the other two evaluation dimensions has no consequence for the attitudes towards surveys. In addition, the association between the subjective burden and survey attitudes is found to be conditional on the subjects' response-latencies. For those who answer the evaluation questions rather fast, the content of these responses has a stronger effect on their generalized attitudes towards surveys. Thus, cognitively more accessible and more salient instances of burdensome interview experiences are especially relevant for the respondents' attitudes. These experiences therefore most likely influence their willingness to participate in future surveys.

1. Einleitung

Eine Reihe empirischer Untersuchungen legt die Vermutung nahe, dass die generalisierte Einstellung gegenüber Umfragen eine wichtige Determinante für die Unterstützung einer konkreten Befragung durch verschiedene Formen kooperativen Verhaltens darstellt. Demnach neigen (potentielle) Befragte mit positiver Umfrageeinstellung stärker zur Teilnahme an Umfragen als Personen mit negativer Umfrageeinstellung (vgl. Goyder 1986). Für die Erklärungskraft dieses Faktors spricht ebenfalls der beobachtete Zusammenhang zwischen der generalisierten Umfrageeinstellung und der für eine erfolgreiche Rekrutierung von Interviewteilnehmern notwendige Aufwand. So erweisen sich spontane Verweigerer, die erst im zweiten Anlauf zu einem Interview bewegt werden konnten, verglichen mit spontanen Teilnehmern als deutlich kritischer gegenüber Umfragen eingestellt (Erbslöh/Koch 1988; Stinchcombe/Jones/Sheatsley 1981). Weitere Belege für den Einfluss der

Umfrageeinstellung auf die Kooperationsbereitschaft der Befragten ergeben sich daraus, dass sich Teilnehmer an einer postalischen Befragungen mit positiver Umfragebewertung stärker an die Fragebogenanweisungen halten, die Fragebögen schneller ausgefüllt zurückschicken, weniger einzelne Fragen unbeantwortet lassen und eher bereit sind, auch in der Zukunft wieder an einer Umfrage teilzunehmen (Rogelberg/Fisher/Maynard/Hakel/Horvath 2001).

Die angesprochenen Ergebnisse zeigen, dass die generelle Haltung der jeweils anvisierten Zielgruppe von Befragten gegenüber Umfragen einen deutlichen Einfluss auf die letztendlich realisierte Datenqualität ausüben kann. Vor diesem Hintergrund stellt sich die Frage, durch welche Prozesse vermittelt und durch welche Faktoren bedingt eine bestimmte generalisierte Bewertung von Umfragen zustande kommt. Diverse Studien aus anderen Themenbereichen lassen schlussfolgern, dass direkte Erfahrungen mit dem jeweiligen Einstellungsobjekt eine zentrale Determinante für die resultierenden Einstellungsinhalte verkörpern (vgl. beispielsweise: Hazzard 1983; Jay/Willis 1992). Nur wenige Untersuchungen hingegen widmen sich der Bedeutsamkeit der Interviewerfahrung für die Entstehung einer bestimmten Einstellung gegenüber Umfragen (vgl. Porst 1998: 4). Dabei gilt das Interesse zumeist der Frage, wie sich unterschiedliche Formen der Belastung durch Meinungsumfragen - beispielsweise durch die Länge der Interviews oder die Häufigkeit der Anfragen für eine Befragungsteilnahme - auf die Bewertung der aktuell durchgeführten Interviews auswirken (Költringer 1992; Sharp/Frankel 1983). Die subjektive Bewertung anderer Aspekte der Interviewerfahrung und deren relative Wirksamkeit für die generalisierten Einstellungen gegenüber Umfragen bleiben hierbei bisher ausgespart.

Die vorliegende Untersuchung schließt an den bisherigen Forschungsstand an und erweitert diesen um eine Analyse der Mehrdimensionalität der Bewertung von Umfrageerfahrungen. Dabei steht einerseits die Frage nach der Differenzierung einzelner Urteilsdimensionen, vor allem aber deren relative Bedeutsamkeit für das Image der Umfrageinstitution im Mittelpunkt. Andererseits werden zusätzlich zu der Analyse dieser generellen Erfahrungsbewertungen auch konkrete und möglicherweise ebenfalls bewertungsrelevante Aspekte vergangener Umfrageteilnahmen berücksichtigt. Hierbei handelt es sich um den Befragungsmodus und die Sponsorenschaft bei in der Vergangenheit absolvierten Befragungen, das insgesamt vorliegende quantitative Ausmaß der Interviewerfahrung, die zeitliche Distanz zu den vorangegangenen Befragungen und die Bedeutung der Täuschung der Befragten durch als Interview getarnte Verkaufsgespräche. Abgesehen von der grundlegenden Zielsetzung einer Erforschung der Determinanten der generalisierten Umfrageeinstellung und der damit verbundenen Kooperationsbereitschaft der Befragten zielt die Analyse auch auf die praxisnahe Fragestellung, ob und in welcher Weise sich die Markt- und Mei-

nungsforschung auf das Umfrageklima in der Bevölkerung auswirkt und sich die Umfrageinstitute somit die Randbedingungen für ihre zukünftige Arbeit selbst schaffen.

2. Forschungsstand

In der Bundesrepublik haben sich in den letzten Jahren die Anteile von Befragten, die ihre Interviewerfahrung als eher angenehm (ca. 50 Prozent) oder als unangenehm (ca. 20 Prozent) beurteilen, kaum verändert (vgl. Forsa 1991, 1993a, 1993b). Zwar nehmen die Bundesbürger im internationalen Vergleich immer noch eine vergleichsweise skeptische Haltung gegenüber Umfragen ein, dennoch lässt sich trotz steigender Umfragetätigkeit keine Umfragemüdigkeit feststellen und die Profession genießt ein insgesamt gutes Image (Forsa 1993a, 1993b; Leiblein/Klass 1990; Leiblein/Oglesby 1993: 48). Die amerikanische Umfrageforschung hingegen konstatierte zumindest in der Vergangenheit steigende Zahlen an Personen, die sich negativ gegenüber Umfragen äußern (Kohut 1986, Schleifer 1986), ohne jedoch von einer Krise der Markt- und Meinungsforschung zu sprechen (Kohut 1986: 8; McDaniel/Verille/Madden 1985: 77).

Als eine der ersten Studien, die sich mit der Wirkung von Umfragen auf die Befragungsteilnehmer und somit quasi mit sich selbst befasste, gilt Sjöbergs „A Questionnaire on Questionnaires“ (1955).² Die Mehrheit der Surveyforscher geht seitdem davon aus, „... that the interview experience is on balance a positive rather than a burdensome one“ (Sharp/Frankel 1983: 43). Es ist jedoch bekannt, dass ein Teil der Befragten bestimmte Items als Eingriff in die Privatsphäre empfindet (vgl. zum Beispiel Hartmann/Isaacson/Jurgell 1968). Da die Einwilligung in ein Interview bereits eine gewisse positive Haltung gegenüber Umfragen impliziert, können keine Rückschlüsse auf die Wahrnehmung von Verweigerern gezogen werden. Es bleibt somit ungeklärt, ob die Gruppe der stetigen Nonresponder Umfragen nicht generell als eine Verletzung ihrer Intimsphäre bewertet.

2.1 Allgemeine Bestimmungsfaktoren der Einstellung gegenüber Umfragen

Generalisierte Einstellungen beinhalten eine gleichförmige und situational unabhängige Bewertung aller in der betreffenden Kategorie zusammengefassten Einstellungsobjekte. Insofern ist die Verwendung von Einstellungen als Urteilsgrundlage und als Kriterium für die Wahl des (kooperativen) Verhaltens eigentlich unvereinbar mit der Basisversion der Theorie rationalen (Befragten-) Verhaltens. Im Rahmen einer erweiterten Variante dieses Erklärungsansatzes lassen sich Einstellungen als Ergebnis der Selektion einer quasi

² Nachfolgende Untersuchungen sind bis heute zumeist Forschungsaktivitäten großer Markt- und Meinungsforschungsinstitute zu verdanken (Goyder 1986: 28).

bewertungsmäßigen „Verfassung“ durch Akteure mit begrenzter Rationalität begreifen (Vanberg 2002). Aus dieser Perspektive treffen die Akteure bei Themen mit geringem Konsequenzenpotential eine grundsätzliche Entscheidung über die Bewertung einer ganzen Klasse von Einstellungsobjekten und die Art des gegenüber diesen Objekten angemessenen Verhaltens. Die Akteure vermeiden so den (kognitiven) Aufwand, der mit einer immer wieder neuen Entscheidung in wiederkehrenden Handlungssituationen verbunden wäre. Da die einstellungsmäßige Grundhaltung aber beim Vorliegen neuer Informationen über die Einstellungsobjekte überprüft wird, ist auch die Selektion einer Einstellung gegenüber Umfragen an einer Kosten-Nutzen-Abwägung orientiert.

So beurteilen die Befragten unter anderem, ob es sich bei Befragungen um eine sinnvolle und akkurat durchgeführte Unternehmung handelt (vgl. Forsa 2000; Roper 1986: 13; Stinchcombe/Jones/Sheatsley 1981). Eine besondere Bedeutung kommt in diesem Zusammenhang dem Umfragethema zu. So bewerten Befragte Umfragen dann positiver und nehmen eine kooperativere Haltung ein, wenn sie an Umfrageergebnissen interessiert sind und diesen einen hohen Unterhaltungswert zumessen (Jones/Sheatsley/Stinchcombe 1979: 69). Das Gleiche trifft auch dann zu, wenn Befragungen und deren Ergebnisse als (gesellschaftlich) relevant erachtet werden (Bradburn 1979). Speziell Letzteres scheint auf die deutsche Bevölkerung in besonderem Maße zuzutreffen, da sich die Beantwortung von Fragen mit einem politischen Inhalt in jüngerer Zeit und im Vergleich zu anderen Ländern einer wachsenden Beliebtheit erfreut (Forsa 1993a; Porst 1998: 10). Analog dazu wird der politischen Meinungsforschung eine hohe Zweckdienlichkeit attestiert (Forsa 1993a). Die Akzeptanz von Umfragen wird auch dadurch verstärkt, dass die Befragten selbst gerne Publikationen von Umfrageergebnissen in der Presse verfolgen.³

Die Umfrageeinstellung spiegelt gleichzeitig die Einschätzung der Akteure hinsichtlich der Belastungen und damit der wahrgenommenen Kosten durch Umfragen wider. Beispielsweise wird hinterfragt, ob in Interviews generell zu persönliche Fragen gestellt werden, ob insgesamt zu viele Umfragen durchgeführt werden (Roper 1986: 14; Sharp/Frankel 1983: 42) oder ob der benötigte Zeitaufwand in einem negativen Verhältnis zu dem erwartbaren Nutzen steht (Roper 1986: 13). Eine Zustimmung zu diesen Überlegungen erhöht den Kosten-Faktor und macht eine negative Umfrageeinstellung wahrscheinlich.

2.2 Bisherige Ergebnisse erfahrungsbasierter Determinanten der Umfrageeinstellung

In der Einstellungsforschung wird gemeinhin davon ausgegangen, dass ausgewählte und

³ Dies gilt vornehmlich für junge und formal gebildete Personen (Forsa 1993a; Leiblein/Oglesby 1993: 50).

gefilterte Erlebnisinhalte aus der Erfahrungswirklichkeit bei der Herausbildung einer Einstellung herangezogen werden (Eagly/Chaiken 1993; Hartfiel/Hillmann 1982; Hazzard 1983; Jay/Willis 1992). Auch hinsichtlich der Einstellung gegenüber Umfragen liegen Hinweise darauf vor, dass diese bei Personen, die bereits zuvor an einer Studie teilgenommen haben, durch die hierbei gemachten Erfahrungen geprägt ist. Konnte persönlicher Nutzen aus der Teilnahme gezogen werden - hat die Befragung zum Beispiel Spaß gemacht oder wurde sie als angenehme Abwechslung empfunden - wirkt sich dies positiv auf die generalisierte Bewertung von Befragungen aus (Porst 1998: 10). Dagegen können negative Erlebnisse mit Umfragen, wie etwa der Versuch unter dem Vorwand einer Meinungsumfrage Zugang zu potentiellen Kunden zu erhalten, zu einer negativen Haltung gegenüber Umfragen führen. Die unterschiedlichen Belastungen bei Befragungsteilnahmen in der Vergangenheit repräsentieren eine weitere Gruppe vermeintlicher Negativfaktoren (Bradburn 1979; Sharp/Frankel 1983). Bradburn beschreibt die Belastung jedoch als ein subjektives Phänomen, welches nicht in der Natur des Interviews begründet liegt, sondern als Reaktion des Befragten auf die besonderen Aspekte der jeweiligen Interviewsituation verstanden werden muss. Sein Ansatz zur Identifizierung relevanter Belastungen für Befragte geht von vier Faktoren aus: der Interviewlänge, dem (kognitiven) Aufwand, den der Befragte bei der Beantwortung der Fragen aufbringen muss, der Häufigkeit, mit der ein Befragter an Untersuchungen teilnimmt und dem psychischen Stress, den heikle oder unangenehme Fragen bei dem Befragten auslösen können (Bradburn 1979).

2.2.1 Erfahrung mit vorgetäuschten Interviews

Eine Determinante für die allgemeine Bewertung von Umfragen kann in dem zwar verbotenen, aber dennoch nachweislich praktizierten Missbrauch eines Interviews zu Verkaufszwecken gesehen werden (Baxter 1964; Schleifer 1986; Sheets/Radlinski/Kohne/Brunner 1974). Rugg fand 1970 bei einer Interviewer-Befragung der Opinion Research Cooperation heraus, dass 51 Prozent der von ihm befragten Interviewer in der Vergangenheit bereits mindestens einmal von Zielpersonen mit der Begründung abgelehnt wurden, dass diese vermuteten, es würde sich um ein getarntes Verkaufsgespräch handeln (Rugg 1971: 625). Untersuchungen über den Anteil von Befragten, die Erfahrungen mit vorgetäuschten Interviews gemacht haben, zeigen ein inkonsistentes Bild. So berichtet Biel (1967), dass bereits 60 Prozent der Befragten solche Erfahrungen gemacht haben, während in einer Studie von Schleifer (1986) lediglich 36 Prozent solcher Fälle dokumentiert werden können. Die tatsächliche Häufigkeit derartiger Missbräuche ist für die aktuelle bundesdeutsche Situation nicht bekannt. Aus Täuschungsversuchen können Irritationen und Verärgerung in der Bevölkerung entstehen, die den Annahmen einiger Methodenforscher zufolge zu Misstrauen gegenüber der Umfrageforschung, zu einer negativen Umfrageeinstellung und damit möglicherweise zu einer erhöhten Verweigerungsrate

führen (McDaniel/Verille/Madden 1985: 79; Schleifer 1986: 22). Ob sich diese Verärgerung in der überdauernden Einstellung zu Umfragen niederschlägt, ist unbekannt (Sheets/Radlinski/Kohne/Brunner 1974: 263).

2.2.2 Interviewlänge

Hinsichtlich des Einflusses der Interviewlänge auf die wahrgenommene Belastung durch Umfragen liegen einige Ergebnisse aus empirischen Studien vor (vgl. u.a. Schleifer 1986, Sharp/Frankel 1983). So bevorzugen nur 14 Prozent der Befragten eine persönlich/mündliche Art der Befragung, während 28 Prozent am liebsten schriftlich und 40 Prozent telefonisch befragt werden möchten (Forsa 2000). Die im Durchschnitt deutlich längere Interviewdauer bei persönlichen Befragungen kann als eine Ursache für diese Präferenz der Befragten gewertet werden. Der Tenor der meisten Untersuchungen lautet so auch, dass sich der Zeitaufwand, der für ein Interview geopfert werden muss, tatsächlich ab einer bestimmten Interviewlänge zuungunsten der Bewertung von Umfragen niederschlägt und möglicherweise dazu führt, dass auch die Bereitschaft einer erneuten Befragungsteilnahme bei den Befragten sinkt (Nederhof 1981: 96). Allerdings wird der negative Effekt besonders langer Interviews relativiert, wenn die betreffende Studie von den Befragten als sozial relevant und damit als wertvoll eingestuft wird (Sharp/Frankel 1983).

2.2.3 Häufigkeit der Befragung

Im Zuge steigender Umfragetätigkeit in Wissenschaft und Marktforschung entwickelte sich die Sorge, eine häufige Befragung könne negative Effekte wie „research tiredness“ oder „over surveying“ verursachen (Dillman/Carpenter/Christenson/Brooks 1974; Groves/ Kahn 1979; Nederhof 1981), welche schließlich, vermittelt durch eine eventuelle Beeinträchtigung der Umfrageeinstellung, in eine sinkende Teilnahmebereitschaft münden. Da keine Untersuchungen existieren, die sich explizit mit dem Einfluss einer häufigen Partizipation an Interviews auf die generalisierte Umfrageeinstellung befassen, wohl aber welche, die sich dem Zusammenhang zwischen Befragungshäufigkeit und Teilnahmebereitschaft widmen, sind wir gezwungen, die Teilnahmebereitschaft als Indikator der Haltung gegenüber Umfragen anzunehmen. Die Ergebnisse dieser Studien fallen widersprüchlich aus. Sharp und Frankel (1983) fanden keine Korrelation zwischen der Befragungshäufigkeit und der Teilnahmebereitschaft, was in unserer Analogie bedeutet, dass keine Effekte hinsichtlich der Umfrageeinstellung auftreten. Nederhof verweist auf eine höhere und schnellere Kooperationsbereitschaft bei Personen mit Interviewerfahrung (Nederhof 1981: 111; Nederhof 1986: 282) sowie eine positivere Haltung zu Umfragen als bei Erstbefragten (Nederhof 1986; vgl. auch McDaniel/Verille/Madden 1985: 78). Es liegen Hinweise darauf vor, dass positive Erfahrungen die Teilnahmebereitschaft erhöhen, negative Erfahrungen sie senken und keine Vorerfahrung mit einer unterdurchschnittlichen

Bereitschaft zur Partizipation gleichzusetzen sind. Der Effekt neutraler Umfrageerfahrungen ist ungeklärt und muss voraussichtlich zwischen den Extremen angesiedelt werden (Nederhof 1987: 426). Unter Berücksichtigung der generalisierten Einstellung zur Umfrageforschung fand Goyder (1986) einen Zusammenhang zwischen negativer Einstellung zu Befragungen, Befragungshäufigkeit und Reduktion der Teilnahmebereitschaft.

2.2.4 Die Bedeutung unterschiedlicher Sponsoren der Befragung

Die Qualität der Sponsorenschaft jener Umfragen, an denen die Befragten in der Vergangenheit teilgenommen haben, kann als eine mögliche weitere Determinante für die generalisierte Umfrageeinstellung angesehen werden. So kann vermutet werden, dass die Befragten Umfragen von Auftraggebern aus dem wissenschaftlichen Bereich ein höheres Ausmaß an Legitimität zuschreiben, verglichen mit solchen von kommerziellen Sponsoren (Orr 1976). Belege für diese Annahme finden sich in einer Reihe von Studien, die entsprechende Unterschiede in der Kooperationsbereitschaft von (potentiellen) Befragten bei Umfragen von universitären und kommerziellen Auftraggebern nachweisen. Es konnte gezeigt werden, dass die Befragten im Falle einer wissenschaftlichen Einrichtung als Sponsor einer postalischen Befragung eher zur Teilnahme neigen und weniger Einzelfragen unbeantwortet lassen, als dies bei einer kommerziellen Sponsorenschaft der Fall ist (Doob/Freedman/Carlsmith 1973; Hawkins 1979; Jones/Lang 1980, 1983; Peterson 1975; für abweichende Ergebnisse vgl.: Albaum 1987; Presser/Blair/Triplett 1992). Weiterhin finden sich Hinweise auf Unterschiede in der Validität der Angaben nach der Art des Auftraggebers. So konnte mithilfe unabhängiger Validierungsdaten gezeigt werden, dass die Befragten bei einer Universität als Sponsor signifikant korrektere Angaben über den Kaufpreis ihres Wohneigentums machen, als dies gegenüber einem kommerziellen Auftraggeber der Fall ist (Jones/Lang 1980). Da die Befragten somit zu einer stärkeren Unterstützung von Umfragen wissenschaftlicher Sponsoren neigen, könnte vermutete werden, dass unter Umständen auftretende negative Erfahrungen mit Umfragen aus diesem Bereich weniger auf die Bewertung der Umfrageinstitution insgesamt generalisiert werden.

3. Empirische Untersuchung

3.1 Stichprobe

Die Studienteilnehmer entstammen einer lokalen, mehrstufigen Zufallsstichprobe der Wohnbevölkerung einer süddeutschen Großstadt. Die Grundgesamtheit der im Sommer 1999 durchgeführten Studie bildeten Bürgerinnen und Bürger mit Lebensmittelpunkt im Stadtgebiet, die zum Befragungszeitpunkt mindestens 18 Jahre alt waren und die deutsche

Staatsbürgerschaft besaßen. Es wurde eine „Random Walk“-Prozedur als Auswahlverfahren verwendet, wobei die Stichprobe durch Mitarbeiter des Forschungsprojektes gezogen und den Interviewern vorgegeben wurde (Adress-Random). Dabei wurden zunächst nach dem Zufallsprinzip im Stadtgebiet Startpunkte bestimmt, von denen ausgehend eine Haushaltsstichprobe generiert werden konnte. Innerhalb der Haushalte wurde immer die Person für eine Befragung ausgewählt, die vor dem Kontaktzeitpunkt als letzte Geburtstag hatte. Mit 139 verwertbaren Interviews ließ sich nur eine relativ schlechte Ausschöpfungsquote von 40,4 Prozent realisieren. Ein Vergleich der Merkmale der im Datensatz repräsentierten Befragten mit der bekannten Sozialstruktur des Auswahlgebietes ergibt die üblicherweise in Umfragedaten vorliegenden Abweichungen.⁴ So sind in der Stichprobe Personen mit Abitur um 11, verheiratete Befragte um 10 und Angestellte um 18 Prozentpunkte überrepräsentiert. Dies führt zu einer Unterrepräsentanz der jeweils anderen Bildungs-, Familien- und Erwerbsstatusgruppen. In Hinblick auf das Einkommen der Befragten tritt der bei Surveys bekannte Mittelschichtbias auf, so dass in der Stichprobe Mitglieder der Grundgesamtheit mit geringem und hohem Einkommen unterrepräsentiert sind. Dagegen liegen hinsichtlich des Alters und dem Grad der Erwerbstätigkeit der Befragten keine systematischen Abweichungen von der Zusammensetzung der Grundgesamtheit vor.

3.2 Vorgehensweise

Die Befragung erfolgte in allen Fällen als computergestütztes Interview bei den Befragten zuhause, wobei zufallsgesteuert der Modus der Datenerhebung variiert wurde. Bei der Hälfte der Befragungen lasen die Interviewer die Fragen laut vor und gaben dann die Antworten der Probanden in den Laptop-Computer ein (CAPI-Modus). Im Gegensatz zu dieser „öffentlichen“ Art der Datenerhebung las die andere Hälfte der Befragten beim „privaten“ CASI-Modus die Fragen selbst vom Bildschirm ab und gab ihre Antworten direkt in den Computer ein. Die Interviewer blieben zwar bei beiden Befragungsmodi anwesend und beantworteten eventuell auftretende Verständnisfragen, zogen sich aber bei CASI-Interviews soweit zurück, dass sie Antwortinhalte nicht wahrnehmen konnten. Durch diese Modus-Unterschiede sollte die Privatheit der Interviewsituation und damit das Vorliegen unterschiedlicher Grade sozialen Drucks operationalisiert werden. Es lässt

⁴ Die Vergleichsdaten für die Bildungsabschlüsse, die Verteilung des Erwerbsstatus und die repräsentative Einkommensverteilung beziehen sich mangels spezifischer Daten für die Grundgesamtheit im Stadtgebiet auf die Bevölkerung Westdeutschlands im April 1996 (Statistisches Bundesamt 1998). Dagegen liegen dem Vergleich hinsichtlich des Ausmaßes der Erwerbstätigkeit Zahlen auf der Ebene des Bundeslandes der Stadtregion vor (Statistisches Landesamt Baden-Württemberg 1998). Die Aussagen über die Repräsentativität der im Datensatz vertretenen Altersstruktur beruht auf spezifischen Zahlen der Stadt, die die Grundgesamtheit definiert (Stadt Mannheim 1998).

sich somit die Möglichkeit testen, dass eventuell beobachtete Zusammenhänge zwischen den Bewertungen der Umfrageerfahrung und den Angaben über die generalisierte Umfrageeinstellung alleine auf das Bestreben der Befragten zur Erfüllung der Norm zu konsistentem Antwortverhalten zurückgehen. Trifft eine solche Alternativklärung zu, so sollten bei der im CAPI-Modus vorliegenden Wahrnehmbarkeit und damit Sanktionierbarkeit des Antwortverhaltens stärkere Zusammenhänge zwischen der Erfahrungsbewertung und der Umfrageeinstellung beobachtbar sein. Die Befragten wurden nach dem Zufallsprinzip entweder der CAPI- oder CASI-Experimentalbedingung zugewiesen. Die Umfrage wurde den Befragten in einem vorab zugestellten Anschreiben als Befragung über „gesellschaftliche Probleme“, über „aktuelle politische Themen“ und über „Erfahrungen bei früheren Umfrageteilnahmen“ angekündigt. Diese Ankündigung war einerseits daher sehr allgemein formuliert, um die Breite der tatsächlich behandelten Befragungsthemen abdecken zu können. Andererseits wurde auch deshalb bewusst nicht stärker auf die Umfrageerfahrung als Schwerpunkt der Befragung eingegangen, um eine Selbstselektion der Befragten nach diesem Merkmal zu vermeiden. Die Interviews dauerten im Durchschnitt 35 Minuten.

3.3 Messinstrumente zur Operationalisierung der verwendeten Konstrukte

Die in diesem Artikel untersuchte Hauptfragestellung erfordert die Operationalisierung der generalisierten Einstellung zu Umfragen, die Erfassung der Bewertung der Umfrageerfahrung durch die Befragten sowie die Messung der kognitiven Zugänglichkeit dieser Bewertungen. Hinsichtlich Letzterer gehen wir von einer Moderatorfunktion dergestalt aus, dass die Erklärungskraft der Bewertungen der vergangenen Interviewerfahrung für die generalisierte Umfrageeinstellung mit der kognitiven Zugänglichkeit dieser Bewertungen zunimmt. Die Zusammenhänge zwischen der Bewertung vergangener Befragungen und der Umfrageeinstellung sollen unter Kontrolle einer Reihe weiterer Aspekte der Umfrageerfahrung getestet werden. So wurde erhoben, ob die Befragten in der Vergangenheit Opfer einer angeblichen Umfrage wurden, bei der in Wirklichkeit Produkte verkauft werden sollten. Außerdem wurden das quantitative Ausmaß der vergangenen Interviewteilnahmen, die zeitliche Distanz dieser Erfahrungen zum aktuellen Messzeitpunkt sowie die Sponsorenschaft und der Befragungsmodus vergangener Umfragen erfasst.

- *Generalisierte Umfrageeinstellung:* Auf der Grundlage vorliegender Studien (Dran/Hildreth 1995; Erbslöh/Koch 1988; Forsa 1993a, 1993b, 1996; Goyder 1986; Porst 1998; Stinchcombe/Jones/Sheatsley 1981) haben wir zuerst 31 Items zur Erfassung der generalisierten Umfrageeinstellung zusammengestellt und in einer Vorstudie einer Zufallsstichprobe vorgelegt. Auf der Grundlage der so realisierten Daten wurden 16 Einstellungsfragen mit einem Höchstmaß an interner Validität ausgewählt (vgl. die Itemformulierungen

in Tabelle 1 unten). Diese 16 Einzelitems bekamen die Befragten in randomisierter Form dargeboten. Die Zustimmung der Befragten zu den positiven und negativen Iteminhalten wurden auf einer siebenstufigen Antwortskala (1 „stimme voll und ganz zu“ bis 7 „lehne voll und ganz ab“) erhoben. Nach der Angleichung der Item-Polung wurde für jeden der Befragten ein Durchschnittswert seiner Einstellungsangaben über Umfragen berechnet. Die resultierende Skala variiert zwischen dem Wert 1 (sehr negative Umfrageeinstellung) und dem Wert 7 (sehr positive Umfrageeinstellung). Die interne Konsistenz der Gesamtskala kann als hinreichend angesehen werden (Cronbachs $\alpha = 0.73$).

- *Bewertung der Umfrageerfahrung*: Zur Bewertung unterschiedlicher Aspekte der letzten Umfrageerfahrung standen den Befragten 6 Einzelitems zur Verfügung. Dabei bezogen sich jeweils zwei Items auf die wahrgenommene Belastung durch die letzte Umfrageteilnahme (zeitlicher Aufwand und notwendige Anstrengungen), auf den bewerteten Umfragenutzen (interessante Erfahrung und Spaß-Komponente) sowie auf das erfahrene Frageverständnis (unklare Formulierung und nicht erkennbarer Sinn; vgl. Tabelle 2 für die Formulierung der Items). Die Zustimmung der Befragten zu den Iteminhalten ist ebenfalls mittels einer siebenstufigen Antwortskala (1 „stimme voll und ganz zu“ bis 7 „lehne voll und ganz ab“) gemessen worden.

- *Kognitive Zugänglichkeit der Bewertungen der Umfrageerfahrung*: Als Indikator für die kognitive Zugänglichkeit der Erfahrungsbewertung werden die durchschnittlichen Antwortlatenzen bei der Beantwortung der Bewertungsfragen herangezogen. Die Validität der Antwortlatenzen als Indikator für die Stärke, für die spontane Aktivierbarkeit und für die Verhaltensrelevanz sowie Stabilität von (Einstellungs-) Urteilen ist durch eine Reihe von Studien bestätigt (Bargh 1996; Bassili 1993; Bassili/Bors 1997; Bassili/Fletcher 1991; Stocké 2002). Die Erfassung der Antwortgeschwindigkeit erfolgte in der vorliegenden Studie im Falle von CAPI-Interviews dergestalt, dass die Interviewer direkt nach dem Vorlesen der Frage die Zeitmessung eingeschaltet und diese bei einer Antwortäußerung der Befragten sofort wieder gestoppt haben. Bei den CASI-Befragungen wird den Teilnehmern der Item-Text zusammen mit den Antwortoptionen auf der gleichen Bildschirmseite präsentiert. Da hier die Zeitmessung beim Aufruf dieser Seite automatisch eingeschaltet und im Moment der Antworteingabe angehalten wird, umfassen die CASI-Latenzen zusätzlich zu der für die Urteilsfindung benötigten Zeit auch die Lesezeit der Itemformulierungen. Die daher bei selbst-administrierten Interviews systematisch längeren Antwortlatenzen wurden korrigiert, indem von jeder individuellen Latenzzeit die durchschnittlich beim betreffenden Item benötigte Lesezeit abgezogen

wurde.⁵ Nach dieser Korrektur liegen keine systematischen Unterschiede in den Antwortgeschwindigkeiten zwischen den Erhebungsmodi vor, wohingegen die Unterschiede in den Latenzen zwischen den Befragten erhalten bleiben.

⁵ Die durchschnittliche Lesezeit pro Item bei CASI-Befragungen berechnet sich, indem die durchschnittliche Antwortlatenz des jeweiligen Items im CAPI-Modus von der CASI-Antwortlatenz desselben Items subtrahiert wird.

Bei der Erfassung von Antwortlatenzen im Kontext einer „normalen“ Bevölkerungsumfrage sind Messstörungen unvermeidlich. So spiegeln die gemessenen Latenzzeiten beispielsweise immer dann nicht allein die zur Antwortgenerierung benötigte Zeit wider, wenn die Befragten Verständnisfragen stellen, während der Beantwortung der Fragen in Interaktion zum Interviewer treten oder während der Antwortgenerierung durch Kontextfaktoren abgelenkt werden (Bassili 1996). Unter diesen Umständen wurden die betroffenen Antwortlatenzen durch die Interviewer als invalide markiert und von der Datenanalyse ausgeschlossen. Dies war in der vorliegenden Studie bei insgesamt 7 Prozent der notwendigen Messungen der Fall. Um eine systematische Selektion der Stichprobe nach den Ursachen der Störungen zu vermeiden, wurden die fehlenden Werte durch Mittelwert-Imputation ergänzt. Die durch die Befragungssoftware bedingte technische Obergrenze der Messgenauigkeit bei der Zeiterfassung liegt bei einem Hundertstel einer Sekunde.

- *Quantitatives Ausmaß der Erfahrung mit Umfragen insgesamt und mit als Interviews getarnten Verkaufsversuchen:* Als Indikator für das vorliegende Ausmaß an Interviewerfahrung mussten die Befragten einerseits angeben, ob sie vor der aktuellen Befragung bereits an einer anderen Umfrage teilgenommen haben. Andererseits wurde erfasst, wie häufig dies insgesamt und in den letzten 12 Monaten der Fall war. Auf der Grundlage dieser Informationen wurde die quantitative Umfrageerfahrung der Befragten bestimmt. Aus älteren amerikanischen Untersuchungen geht hervor, dass eine Täuschung von Befragten durch als Umfrage getarnte Verkaufsgespräche deren Kooperationsbereitschaft in der Zukunft und damit möglicherweise ihre Haltung gegenüber Umfragen generell negativ beeinflusst (siehe Kap. 2.2.1). Um entsprechende Einflüsse derartiger Erfahrungen auf die generalisierte Umfrageeinstellung überprüfen zu können, wurden die Studienteilnehmer ebenfalls danach befragt, ob sie bei ihren Umfrageteilnahmen in der Vergangenheit solchen Täuschungsversuchen begegnet sind.

- *Sponsorenschaft, Befragungsmodus und zeitliche Distanz der letzten Interviewteilnahme:* Bei Befragten mit Interviewerfahrung wurde die Sponsorenschaft der letzten Umfrage erfasst. Hierbei mussten die Probanden erinnern, ob diese Befragung durch ein Marktforschungsinstitut, durch eine wissenschaftliche Institution oder durch einen anderen Sponsor in Auftrag gegeben wurde.⁶ Ferner wollten wir hinsichtlich des Befra-

⁶ In vielen Umfragen unterscheidet sich der Auftraggeber von der die Befragungen durchführenden Organisation. Unter diesen Umständen werden im Ankündigungsschreiben und in der Kontaktphase oft beide und damit möglicherweise unterschiedliche Sponsorentypen genannt. Die vorliegende Operationalisierung bezieht sich somit auf die subjektive Zuschreibung der vergangenen Interviewerfahrung zu einer der Sponsorentypen. Diese Wahrnehmung durch die Befragten muss nicht mit dem objektiven Auftraggeber der Befragung übereinstimmen.

gungsmodus erfahren, ob es sich bei der letzten Umfrageteilnahme um eine face-to-face Befragung zuhause, um ein Telephoninterview, um eine schriftliche Befragung oder um eine Befragung auf der Straße gehandelt hat. Da getestet werden soll, ob eventuell beobachtete Einflüsse der subjektiven Bewertung der letzten Interviewerfahrung mit der zeitlichen Distanz dieser Erfahrung abgeschwächt werden, haben wir zusätzlich das Jahr und den Monat der unmittelbar dem aktuellen Interview vorangegangenen Umfrageteilnahme erhoben.

3.4 Empirische Ergebnisse

Die Darstellung der empirischen Ergebnisse gliedert sich in drei Abschnitte. Im ersten Teil werden die realisierten Daten hinsichtlich der Umfrageeinstellung, der Bewertung der letzten Umfrageerfahrung und der anderen relevanten Erklärungsfaktoren deskriptiv dargestellt. Im zweiten Abschnitt folgt die Analyse, ob sich die allgemeine Umfragebewertung nach der sozialstrukturellen Lage der Befragten, dem Ausmaß ihrer Umfrageerfahrung sowie der Sponsorenschaft und dem Befragungsmodus der letzten Umfrageteilnahme unterscheidet. Im dritten und wichtigsten Teil der Analyse können wir zeigen, ob und in welcher relativen Stärke sich verschiedene Bewertungsdimensionen der vergangenen Interviewerfahrung auf die Umfrageeinstellung der Befragten ausgewirkt haben.

3.4.1 Deskriptive Ergebnisse

Im folgenden Abschnitt sind die Angaben der Befragten über ihre generalisierten Einstellungen gegenüber Umfragen, über die Bewertung der letzten Umfrageteilnahme sowie über die anderen möglicherweise relevanten Erklärungsfaktoren dargestellt.

- *Generalisierte Einstellung gegenüber Umfragen:* Die Umfrageeinstellung erweist sich insgesamt als relativ positiv (vgl. Tabelle 1). Abgesehen von einem Item liegt die durchschnittliche Bewertung hinsichtlich aller abgefragter Items über dem Mittelpunkt der Antwortskala und damit im positiven Einstellungsbereich. Dabei wird mit einem durchschnittlichen Skalenwert von 5.4 die Nützlichkeit von Umfragen für die gesellschaftlichen Kernbereiche Wirtschaft, Politik und Wissenschaft am positivsten eingeschätzt (Item 1). Die Belastung durch zu viele Fragen in Interviews wird mit einer mittleren Bewertung von 3.9 noch am kritischsten beurteilt (Item 15). Obwohl der Entscheidung zur Teilnahme an einer Befragung eine Filterfunktion hinsichtlich der repräsentierten Umfrageeinstellungen zukommt (Erbslöh/Koch 1988), findet sich in der Stichprobe ein substantielles Ausmaß an Heterogenität zwischen den Befragten. So variieren die durchschnittlichen Antworten der Befragten auf der Gesamtskala zur Erfassung der Umfrageeinstellung zwischen einem relativ negativen Wert von 2.8 und einer sehr positiven Bewertung von 6.8.

Tabelle 1: Items zur Erfassung der generalisierten Einstellung gegenüber Umfragen

Itemformulierung	Mittelwert (Std.)
1. Umfragen sind wichtig für Wissenschaft, Politik und Wirtschaft	5.44 (1.56)
2. Meinungsumfragen machen die Gesellschaft demokratischer	4.37 (1.83)
3. Die Teilnahme an Umfragen ist in meinem eigenen Interesse	4.60 (1.83)
4. Es sollte für jeden selbstverständlich sein, an einer Umfrage teilzunehmen	4.66 (1.97)
5. Umfragen bringen Abwechslung und sind interessant	4.50 (1.73)
6. Bei Umfragen kann man endlich mal seine Meinung sagen	4.89 (1.92)
7. Die Ergebnisse von Umfragen sind in den meisten Fällen richtig	4.73 (1.53)
8. Die meisten Umfragen sind seriös und werden verantwortungsvoll durchgeführt	4.70 (1.53)
9. Bei Umfragen werden normale und repräsentative Menschen befragt	4.86 (1.83)
10. Die Teilnehmer an Umfragen sagen meistens ihre wahre Meinung	5.01 (1.74)
11. Die Teilnehmer an Umfragen geben sich sehr viel Mühe, richtig zu antworten	5.30 (1.47)
12. Durch Umfragen wird man nur von wichtigeren Dingen abgehalten	4.70 (1.81)
13. Bei Umfragen wird häufig etwas gefragt, was niemanden etwas angeht	4.02 (1.87)
14. In Deutschland werden viel zu viele Umfragen durchgeführt	4.17 (1.90)
15. Es ist anstrengend in Umfragen so viele Fragen zu beantworten	3.89 (2.01)
16. Interviewer sind unangenehme und aufdringliche Leute	4.77 (1.92)
Gesamtskala	4.66 (0.79)
Fallzahl	N=139

Skalenwert 1 = negative Umfrageeinstellung; Skalenwert 7 = positive Umfrageeinstellung. Die Angaben bei den Items 1 bis 11 gehen in einer rekodierten Form in die Analyse ein.

- *Bewertung, Dimensionalität und kognitive Zugänglichkeit der Umfrageerfahrung:* Auf der Grundlage der Ergebnisse einer Hauptkomponentenanalyse lässt sich feststellen, dass die Gesamtbewertung der letzten Umfrageerfahrung auf drei unabhängigen, latenten Dimensionen beruht (vgl. die Ergebnisse in Tabelle 7 im Anhang). Jeweils zwei der insgesamt sechs Items bilden eine gemeinsame latente Bewertungsdimension. Die drei Bewertungsdimensionen erfassen das Ausmaß der erlittenen Belastungen, den erlebten intrinsischen Nutzen und die wahrgenommenen Defizite in der Verständlichkeit der Fragen bei der letzten Interviewteilnahme (vgl. Tabelle 2).

Tabelle 2: Bewertung unterschiedlicher Aspekte der letzten Umfrageerfahrung und Antwortgeschwindigkeiten bei der Beantwortung der Bewertungsfragen

	Antwortinhalte ¹		Antwortgeschwindigkeit ²	
	Werte- bereich	Mittelwert (Std.)	Werte- bereich	Mittelwert (Std.)
Belastung				
1. Die Befragung hat zu lange gedauert	1/7	4.57 (2.28)	66/1725	330.58 (248.16)
2. Die Befragung war anstrengend	1/7	5.32 (1.85)	66/1093	288.30 (190.54)
<i>Gesamt</i>	<i>1/7</i>	<i>4.95 (1.69)</i>	<i>74/1354</i>	<i>309.44 (180.33)</i>
Nutzen				
3. Die Umfrage war interessant	1/7	4.68 (2.05)	55/4652	361.87 (516.60)
4. Die Befragung hat Spaß gemacht	1/7	4.37 (2.00)	49/2582	269.56 (292.86)
<i>Gesamt</i>	<i>1/7</i>	<i>4.53 (1.83)</i>	<i>74/2656</i>	<i>315.71 (328.53)</i>
Verständnis				
5. Die Fragen waren teilweise unklar formuliert	1/7	4.94 (2.07)	49/1950	366.74 (304.97)
6. Ich habe manchmal den Sinn der Fragen nicht erkennen können	1/7	4.57 (2.26)	60/1510	354.60 (290.79)
<i>Gesamt</i>	<i>1/7</i>	<i>4.76 (1.86)</i>	<i>63/1266</i>	<i>360.67 (235.31)</i>
Gesamtbewertung der Erfahrung	1/7	4.74 (1.22)	126/1141	328.61 (182.11)
Fallzahl	N=91		N=91	

¹ Skalenwert 1=negative Erfahrung; Skalenwert 7=positive Erfahrung; die negativ gepolten Items 1, 2, 5 und 6 wurden entsprechend umkodiert. ² Die Maßeinheit beträgt ein hundertstel einer Sekunde.

Mit einem durchschnittlichen Skalenwert von 4.7 über alle Bewertungsfragen hinweg, sehen die Befragten ihre vergangene Interviewteilnahme als eher positive Erfahrung an. Die Urteile divergieren allerdings zwischen den drei Bewertungsdimensionen, wobei die durchschnittlichen Bewertungen hinsichtlich der subjektiv wahrgenommenen Belastungen mit einem Skalenwert von 4.9 am positivsten und der Unterhaltungswert mit einer mittleren Bewertung von 4.5 am negativsten beurteilt wird. Dieser Mittelwertsunterschied ist marginal signifikant ($t=1.77$, $df\ 90$, $p < 0.1$). Dagegen nimmt die durchschnittliche Bewertung der Frageverständlichkeit beim letzten Interview einen Wert von 4.8 an und unterscheidet sich nicht signifikant von den Urteilen auf den beiden anderen Teilskalen. Die Tatsache, dass insgesamt sehr wenige Befragte außerordentlich negative Umfrageerfahrungen schildern, stellt uns vor dieselbe Frage wie Nederhof (1987: 430): „This may either mean that very few surveys give rise to very unpleasant experiences, or that only

very unpleasant experience is enough reason to turn somebody into a nonresponder for a long time“. Bei den folgenden Analysen werden die Faktoren-Werte der einzelnen Befragten auf jeder der drei Dimensionen der Erfahrungsbewertung herangezogen. Die Latenzen bei der Beantwortung der beiden Bewertungsfragen jeder der drei Dimensionen bilden ebenso jeweils einen singulären Faktor. Auch in diesem Fall werden die Faktoren-Werte als Operationalisierung der kognitiven Zugänglichkeit der Urteile herangezogen.

Die in Tabelle 2 dargestellten Daten zeigen auch, dass die Probanden mit einer durchschnittlichen Antwortzeit von 3.09 Sekunden die Fragen über die belastenden Aspekte ihrer letzten Umfrageteilnahme am schnellsten beantwortet haben. Dagegen erweisen sich die Urteile über die Verständlichkeit des Fragebogeninstrumentes mit einer mittleren Antwortgeschwindigkeit von 3.61 Sekunden als signifikant langsamer ($t=1.99$, $df 90$, $p < 0.05$). Die kognitive Verfügbarkeit der Urteile hinsichtlich des intrinsischen Nutzens der Befragungsteilnahme nimmt mit einem Wert von 3.16 Sekunden eine mittlere Position ein. Die Reaktionszeitenmessungen lassen darauf schließen, dass den Befragten das Ausmaß der Belastungen bei der letzten Interviewteilnahme, im Vergleich zu den anderen Erfahrungsbewertungen, am stärksten salient ist, was als erster Hinweis für deren Bedeutung für die Umfrageeinstellung gewertet werden kann.

- *Quantitative Umfrageerfahrung und Dauer seit der letzten Befragungsteilnahme*: Insgesamt weisen die Studienteilnehmer ein hohes Ausmaß quantitativer Umfrageerfahrung auf. 66 Prozent von ihnen nahmen bereits mindestens einmal zuvor an einer Befragung teil, während im Bundesdurchschnitt am Anfang des Jahres 2000 lediglich 47 Prozent der Bürger in West- sowie 38 Prozent in Ostdeutschland über Umfrageerfahrung verfügte (Forsa 2000). Informanten mit Umfrageerfahrung haben vor der aktuellen Befragung durchschnittlich bereits an 4.0 Interviews teilgenommen. Werden auch Probanden ohne Umfrageerfahrung einbezogen, so liegt dieser Wert bei 2.6 Befragungen. Dabei fanden im Mittel 0.6 Befragungen in den letzten 12 Monaten und 2.0 in der Zeit zuvor statt. Von den Befragten mit Umfrageerfahrung nahmen 28 Prozent bisher an einem oder zwei Interviews teil. Nur 6.4 Prozent der Befragten partizipierten bereits an 10-20 Interviews. Die durchschnittliche Dauer seit der letzten Umfrage liegt bei knapp 24 Monaten.

Tabelle 3: Quantitative Umfrageerfahrung und Dauer seit der letzten Umfrageerfahrung

	Wertebereich	Mittelwert (Std.)	
Anzahl Umfragen in den letzten 12 Monaten	0 / 10	0.60	(1.16)
Anzahl Umfragen älter als 12 Monate	0 / 18	2.04	(2.87)
Umfrageerfahrung gesamt	0 / 20	2.64	(3.59)
Dauer seit letzter Umfrageteilnahme (in Monaten) ¹	1 / 120.5	23.78	(28.22)
Fallzahl	N=139		

¹ nur Befragte mit Umfrageerfahrung, N=91

Tabelle 4: Sponsor und Befragungsmodus der letzten Umfrage sowie Ausmaß der Erfahrung mit vorgetäuschten Interviews (nur Befragte mit Interviewerfahrung)

Sponsor der letzten Umfrage	N	Prozent
Marktforschungsinstitut	56	61.5
Staatliche und wissenschaftliche Institutionen	19	20.9
Anderer/keine Angabe	16	17.6
<i>Gesamt</i>	91	100
Befragungsmodus der letzten Umfrage	N	Prozent
Telefoninterview	37	40.7
Postalische Befragung	11	12.1
Persönliches Interview	15	16.5
Passantenbefragung	23	25.3
Anderer/keine Angabe	5	5.5
<i>Gesamt</i>	91	100
Täuschungen	N	Prozent
Erfahrung mit als Interviews getarnten Verkaufsversuchen	37	40.7
Keine solche Erfahrung	54	59.3
<i>Gesamt</i>	91	100

- *Sponsorenschaft und Befragungsmodus der letzten Umfrageerfahrung*: Im Hinblick auf die Auftraggeber der in der Vergangenheit absolvierten Umfragen spiegelt sich ein erwartetes Bild wider (Tabelle 4). Eine absolute Mehrheit von 61.5 Prozent der Befragten gibt an, dass sie zuletzt an einer kommerziellen Marktforschungsstudie beteiligt war, während nur 20.9 Prozent von staatlichen oder wissenschaftlichen Institutionen interviewt wurden. Den hierbei am häufigsten angewendeten Befragungsmodus stellt mit 40.7

Prozent das Telefoninterview dar, gefolgt von Befragungen auf der Straße (25.3 Prozent), face-to-face Interviews in der eigenen Wohnung (16.5 Prozent) und postalischen Befragungen (12.1 Prozent).⁷ Ein hoher Anteil von 40.7 Prozent der Befragten musste bereits mindestens einmal Erfahrungen damit machen, dass ihre Bereitschaft zu einer Befragungsteilnahme durch den Versuch, Produkte zu verkaufen, missbraucht wurde.

3.4.2 Bedeutung der sozialstrukturellen Gruppenzugehörigkeit und der quantitativen Interviewerfahrung für die Umfrageeinstellung

Bei der Überprüfung sozialstruktureller Unterschiede hinsichtlich der generalisierten Einstellung zu Umfragen lassen sich vier bedeutsame Dimensionen identifizieren (vgl. Tabelle 5, Modell 1).⁸ So bewerten Befragte mit einem festem Lebenspartner Umfragen signifikant positiver als Personen, die nicht in einer Partnerschaft leben. Weiterhin hat die Konfessionszugehörigkeit insofern Einfluss auf die Umfrageeinstellung, als Katholiken Befragungen signifikant negativer bewerten als Angehörige anderer Konfessionen oder konfessionslose Befragte. Dagegen wirkt sich ein Minimum an Einbindung der Befragten in eine Kirchengemeinde – bei gleichzeitiger Kontrolle der Konfession der Befragten – positiv auf die Bewertung von Umfragen aus. So berichtet das Drittel der Befragte, das nie einen Gottesdienst besucht, eine signifikant negativere Umfrageeinstellung verglichen mit der Komplementärgruppe. Die Gegenüberstellung weniger extremer Subgruppen bezüglich der Kirchengangshäufigkeit zeigt hingegen keine statistisch signifikante Differenzierung hinsichtlich der Umfrageeinstellung. Weiterhin verfügen Frauen über eine marginal signifikant positivere Einstellung zu Umfragen als Männer. Für diese Unterschiede in der Umfrageeinstellung lassen sich beim derzeitigen Forschungsstand keine fundierten Erklärungen angeben. Sie werden hier dennoch dokumentiert, weil sich auf deren Grundlage Erwartungen über die unterschiedliche Kooperationsneigung der verschiedenen Sozialgruppen ergeben.

⁷ Sowohl die Verteilung der Sponsoren bei Umfragen als auch die Verteilung der Befragungsmodi spiegeln die Größenordnung der bundesweiten Verhältnisse wider. Laut dem Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute (ADM) wurden in den derzeit 45 Mitgliedsinstituten im Jahr 2000 41 Prozent der Interviews telefonisch durchgeführt, 34 Prozent face-to-face und 22 Prozent schriftlich (vgl. Tabelle 8: "Quantitative Interviews der Mitgliedsinstitute des ADM nach Befragungsart", in: www.adm-ev.de/homepage.html; Subfolder „Zahlen“).

⁸ Abgesehen von den dargestellten signifikanten Einflussvariablen wurde auch die Bedeutung des Alters und Einkommens der Befragten, sowie die des Erwerbsstatus und des Familienstandes überprüft. Da sich die Bewertung von Umfragen nicht nach diesen Dimensionen unterscheidet, werden die Ergebnisse dieser Analysen hier nicht berichtet.

Tabelle 5: Einfluss der sozialstrukturellen Lage und der quantitativen Umfrageerfahrung auf die Einstellung zu Umfragen (OLS-Regressionsergebnisse)

	Modell 1		Modell 2	
	B	(t-Wert)	B	(t-Wert)
Sozialstrukturelle Faktoren				
Hat festen Lebenspartner ¹	.364	(2.46)**	.339	(2.31)**
Römisch katholisch ²	-.324	(2.24)**	-.341	(2.35)**
Besucht nie Gottesdienst ³	-.336	(2.23)**	-.363	(2.41)**
Frau ⁴	.227	(1.75)*	.254	(1.96)*
Quantitative Umfrageerfahrung ⁵				
Noch nie vorher an einer Umfrage teilgenommen	--	--	-.192	(1.23)
Mindestens eine Teilnahme, aber älter als 1 Jahr	--	--	-.381	(2.35)**
Mindestens eine Teilnahme im letzten Jahr	--	--	-.356	(0.90)
R ²	0.10		0.14	
N	139		139	

Signifikanz: * = $p \leq 0.1$; ** = $p \leq 0.05$

Referenzkategorie: ¹ Ohne Partner; ² Andere Konfession; ³ Besucht Gottesdienst zumindest einmal im Jahr; ⁴ Mann; ⁵ Hat im letzten Jahr und zuvor an mindestens einer Umfrage teilgenommen.

Der Zusammenhang zwischen der Häufigkeit der Teilnahme an Umfragen in der Vergangenheit und der Einstellung zu Umfragen gestaltet sich unseren Ergebnissen zufolge komplex (vgl. Tabelle 5; Modell 2). So konnten keine linearen Einflüsse der Gesamthäufigkeit der Umfrageteilnahme, der Teilnahmehäufigkeit in den letzten 12 Monaten oder der Anzahl der Erfahrungen mit Befragungen, die länger als 12 Monate zurück liegen, auf die generalisierte Umfrageeinstellung festgestellt werden. Dies ist allerdings bei einer kombinierten Betrachtung des Vorliegens und der zeitlichen Kontinuität von Umfrageerfahrungen der Fall.⁹ Im Vergleich zu den Befragten, die sowohl in der jüngsten wie auch in der weiter zurückliegenden Vergangenheit kontinuierlich an Umfragen teilgenommen haben (Referenzkategorie), bewerten alle anderen Befragtengruppen Umfragen negativer. Dies trifft insbesondere für diejenigen Personen zu, die entweder nur vor längerer Zeit oder aber nur in jüngster Vergangenheit an Umfragen teilgenommen haben. Befragte ohne jegliche Vorerfah-

⁹ Die erste Kategorie der erklärenden Variable markiert Befragte ohne jegliche Umfrageerfahrung (33.8 Prozent) und die zweite Kategorie Probanden mit mindestens einer Umfrageteilnahme, die aber älter als 12 Monate ist (28.8 Prozent). Die dritte Kategorie beinhaltet Befragte mit Erfahrungen, die ausschließlich im letzten Jahr stattgefunden haben (2.9 Prozent) und Kategorie vier solche Befragte, die an Interviews sowohl in jüngster wie auch in weiter entfernter Vergangenheit teilgenommen haben (34.5 Prozent).

rung hingegen sind im Vergleich zu diesen beiden Gruppen positiver und im Vergleich zu den kontinuierlichen Befragungsteilnehmern negativer gegenüber Umfragen eingestellt (vgl. hierzu auch McDaniel/Verille/Madden 1985: 78). Die vorliegenden Ergebnisse können am ehesten als kombiniertes Ergebnis zweier Prozesse mit unterschiedlicher Kausalrichtung interpretiert werden. Einerseits kann die positive Umfrageeinstellung als Ursache und nicht als Resultat einer kontinuierlichen Teilnahmebiographie verstanden werden. In dieser Gruppe herrscht eine zeitlich stabile und stark verankerte Disposition zur Unterstützung von Umfragen und damit zur Interviewteilnahme. Dagegen könnte vermutet werden, dass der Bruch in der Teilnahmebiographie von Befragten, die zwar früher aber nicht in jüngster Vergangenheit an Umfragen teilgenommen haben, durch schlechte Erfahrungen bewirkt wurde. In der Tat verfügen Befragte mit kontinuierlicher Umfrageteilnahme in der Vergangenheit über eine durchschnittlich positivere Gesamtbewertung ihrer letzten Interviewerfahrung als solche ohne kontinuierliche Befragungsteilnahme. Dieser Bewertungsunterschied zwischen 4.82 und 4.69 ist jedoch nicht statistisch signifikant. Da erst die direkte Erfahrung mit dem Einstellungsgegenstand zu einer Kristallisation und damit Polarisierung der Einstellungen führt, wird die bewertungsmäßige Mittelposition der Befragte ohne bisherige Erfahrung mit Interviews verständlich. Die relativ negative generalisierte Einstellung zu Umfragen von Befragten mit ausschließlicher Interviewerfahrung in jüngster Vergangenheit erscheint uns dagegen schwer interpretierbar. Da es sich hierbei um einen nicht signifikanten Effekt handelt und dieser nur auf knapp drei Prozent der Befragten beruht, wäre eine Replikation dieser Einstellungsunterschiede abzuwarten.

3.4.3 Bedeutung des Befragungsmodus und der Sponsorenschaft der letzten Umfrage

Im Rahmen der in Tabelle 6 dargestellten Regressionsmodelle werden unterschiedliche qualitative Aspekte der vergangenen Interviewerfahrung auf ihre Erklärungskraft für die generalisierte Umfrageeinstellung überprüft. Dabei werden immer die in Tabelle 5 dargestellten sozialstrukturellen Variablen und die quantitative Umfrageerfahrung statistisch kontrolliert. Hierbei auftretende Koeffizienten berichten wir jedoch nicht mehr. Modell 1 überprüft zunächst die Bedeutung des Befragungsmodus und der Sponsorenschaft der letzten Umfrageteilnahme für die aktuell berichtete Umfrageeinstellung. Dabei zeigt sich, dass alle Befragungsmodi der letzten Umfrage in Relation zur face-to-face Befragung zuhause (Referenzkategorie) einen positiveren Einfluss auf die Einstellung zu Umfragen ausüben. Ein Telefoninterview als letzte Befragungsart wirkt sich dabei signifikant und eine Befragung auf der Straße marginal signifikant positiver auf die Umfrageeinstellung aus. Diese Ergebnisse stehen zunächst im Widerspruch zu den von Forsa erhobenen Daten. Hier empfanden die meisten Befragten speziell das persönlich-mündliche Interview in der eigenen Wohnung als angenehm (Forsa 1991, 1993a, 1993b; vgl. auch Reuband

1998: 61). Auf die Frage nach der bevorzugten Erhebungsform sprachen sich jedoch mehr Befragte für das telefonische als für das face-to-face Interview aus (Forsa 2000). Zwei kontradiktorische Einflussgrößen kommen damit bei der Bewertung der Befragungsmethode zum Tragen. Während sich die Telefonumfrage für den Befragten bequemer gestaltet (niedrige Kosten) liegt jedoch auch der damit verbundene Unterhaltungswert im Vergleich zum face-to-face Kontakt niedriger (niedriger Nutzen). Falls zusätzliche Faktoren wie die Anonymität eine wachsende Bedeutung erhalten, wird die Unterhaltsamkeit immer seltener das entscheidende Moment für die Bewertung von Umfragen und damit potentiell für die Teilnahmeentscheidung repräsentieren (vgl. Leiblein/Oglesby 1993: 52). Bei der weiteren Analyse unserer Daten ergeben sich jedoch Hinweise auf die Angemessenheit einer dritten Erklärung der beobachteten Modus-Effekte. Demnach ist die typischerweise größere Länge von face-to-face Interviews der Grund für den hier festgestellten negativen Einfluss dieser Befragungsart (vgl. Abschnitt 4.4). Der Sponsor der letzten Befragung wirkt sich dagegen nicht signifikant auf die generalisierte Umfrageeinstellung aus, so dass kein negativer Einfluss kommerzieller Studien auf das Image von Umfragen feststellbar ist (vgl. dazu auch Forsa 1996, McDaniel/Verille/Madden 1985: 78).

3.4.4 Einfluss vorgetäuschter Interviews und der qualitativen Bewertung der vergangenen Befragungsteilnahme

Die in Tabelle 6 dargestellten Ergebnisse von Regressionsmodell 2 verdeutlichen erstens, dass sich die Erfahrung mit vorgetäuschten Befragungen als „Türöffner“ für Verkaufsgespräche zwar negativ auf die generalisierte Umfrageeinstellung auswirkt, dieser Effekt aber keineswegs ein statistisch abgesichertes Niveau erreicht. Entgegen der in der Methodenforschung oft getroffenen Annahme werden demnach solche Vorfälle nicht der Institution der Umfrageforschung angelastet und richten daher nur wenig Schaden in bezug auf deren Image an.

Als ein zweites und für unsere Untersuchung zentrales Ergebnis lässt sich festhalten, dass die generalisierte Umfrageeinstellung der Befragten immer dann zunehmend positiver ausfällt, wenn diese ihre letzte Umfrageteilnahme als wenig belastend beurteilen, diesem Interview einen größeren Unterhaltungswert zuschreiben und die bei der vergangenen Befragung verwendeten Frageformulierungen als wenig verwirrend ansehen (vgl. Tabelle 6, Modell 2). Dabei wirken sich allerdings ausschließlich die wahrgenommenen Belastungen bei diesem Interview statistisch abgesichert auf die generelle Beurteilung von Befragungen aus. Diese hervorgehobene Erklärungskraft der Belastungsdimension stimmt mit der bereits bei der deskriptiven Beschreibung der Antwortlatenzen festgestellten, relativ zu den anderen Beurteilungsdimensionen stärksten kognitiven Zugänglichkeit dieser Bewertungsdimension überein.

Tabelle 6: Bedeutung unterschiedlicher Bewertungsdimensionen der Umfrageerfahrung für die generalisierte Umfrageeinstellung der Befragten (OLS-Regressionsergebnisse für Befragte mit Umfrageerfahrung)

	Modell 1		Modell 2		Modell 3		Modell 4	
	B	(t-Wert)	B	(t-Wert)	B	(t-Wert)	B	(t-Wert)
Befragungsart ¹								
Telefoninterview	.584	(2.14)**	.453	(1.83)*	.466	(1.78)*	.395	(1.55)
Postalische Befragung	.441	(1.35)	.293	(0.92)	.246	(0.73)	.175	(0.57)
Befragung auf der Straße	.508	(1.79)*	.252	(0.91)	.226	(0.78)	.161	(0.59)
Andere Befragungsart	.176	(0.42)	-.067	(0.16)	.019	(0.04)	-.045	(0.11)
Sponsor der letzten Umfrage ²								
Marktforschungsinstitut	.241	(1.05)	--	--	--	--	--	--
Staatliche oder wissenschaftliche Institution	.357	(1.27)	--	--	--	--	--	--
Hat vorgetäuschte Befragungen erlebt ³								
	--	--	-.248	(1.32)	--	--	--	--
Qualitative Erfahrung bei letzter Umfrage								
Belastung	--	--	.167	(1.96)**	.173	(1.40)	.216	(2.45)**
Nutzen	--	--	.110	(1.29)	.122	(0.91)	.143	(1.74)*
Verständnis	--	--	.077	(0.91)	.028	(0.21)	.124	(1.50)
Moderatorvariable Öffentlichkeit								
Privater Befragungsmodus (CASI) ⁴	--	--	--	--	-.214	(1.26)	--	--
Privat • Belastung	--	--	--	--	-.002	(0.01)	--	--
Privat • Nutzen	--	--	--	--	.009	(0.05)	--	--
Privat • Verständnis	--	--	--	--	.126	(0.72)	--	--

(Fortsetzung auf nächster Seite)

(Fortsetzung von Tabelle 6)

	Modell 1		Modell 2		Modell 3		Modell 4	
	B	(t-Wert)	B	(t-Wert)	B	(t-Wert)	B	(t-Wert)
Moderatorvariable zeitliche Distanz								
Dauer seit der letzten Umfrageteilnahme	--	--	--	--	--	--	.077	(0.83)
Dauer • Belastung	--	--	--	--	--	--	-.086	(1.00)
Dauer • Nutzen	--	--	--	--	--	--	-.039	(0.55)
Dauer • Verständnis	--	--	--	--	--	--	.005	(0.08)
Moderatorvariable Antwortlatenz								
Reaktionszeit (Belastung)	--	--	--	--	--	--	-.131	(1.40)
Reaktionszeit (Nutzen)	--	--	--	--	--	--	-.006	(0.09)
Reaktionszeit (Verständnis)	--	--	--	--	--	--	.144	(1.73)*
Reaktionszeit (Belastung) • Belastung	--	--	--	--	--	--	-.178	(3.20)**
Reaktionszeit (Nutzen) • Nutzen	--	--	--	--	--	--	.075	(1.53)
Reaktionszeit (Verständnis) • Verständnis	--	--	--	--	--	--	.085	(0.93)
Konstante	3.65	(9.95)	4.22	(12.95)	4.18	(12.54)	4.11	(14.00)
R ²	0.22		0.28		0.28		0.46	
N	91		91		91		91	

Signifikanz: * = $p \leq 0.1$; ** = $p \leq 0.05$ Referenzkategorie: ¹ Interview bei den Befragten zuhause; ² Andere oder unbekannte Sponsorschaft; ³ Ohne Erfahrung mit getäuschten Umfragen; ⁴ Öffentlicher Befragungsmodus (CAPI).

Die Ergebnisse von Modell 2 zeigen gleichfalls, dass die Erklärungskraft der unterschiedlichen Erhebungsmodi bei der letzten Umfrageteilnahme durch die allgemeine Bewertung dieser Erfahrung auf den drei Urteilsdimensionen und insbesondere durch die signifikante Bedeutung der Belastungswahrnehmung absorbiert wird. Insofern können wir begründet vermuten, dass der anfänglich beobachtete negative Einfluss von face-to-face Interviews auf die generalisierte Umfrageeinstellung durch die typischerweise größere Länge und damit höhere Belastung der Befragten bei diesem Befragungsmodus vermittelt wird.

3.4.5 Die Privatheit der Antwortsituation als Moderatorvariable

In Modell 3 wird zusätzlich zu den Erfahrungsbewertungen der Befragten die Privatheit des jeweils verwendeten Befragungsmodus berücksichtigt.¹⁰ Hierbei soll überprüft werden, ob die festgestellten Zusammenhänge zwischen der Umfrageerfahrung und der Umfrageeinstellung möglicherweise als Artefakt der Erfassung beider Angaben im gleichen Interview angesehen werden müssen. So könnte vermutet werden, dass sich die Angaben hinsichtlich der zuerst abgefragten Umfrageeinstellung auf die später erfasste Bewertung der Umfrageerfahrung auswirken, weil sich die Befragten zu konsistentem Antwortverhalten gedrängt fühlen und somit bei positiven Einstellungsangaben auch positive Erfahrungsbewertungen berichten. Wenn eine solche Erklärung der beobachteten Zusammenhänge durch die Wirksamkeit von Konsistenzdruck zutrifft, sollte sich jedoch bei CASI- und CAPI-Interviews eine unterschiedlich starke Übereinstimmung zwischen der Erfahrungsbewertung und den Einstellungsangaben finden lassen. Da dieser Druck bei Wahrnehmbarkeit des Antwortverhaltens durch die Interviewer im CAPI-Modus stärker sein sollte, wäre hier mit stärkeren Zusammenhängen zu rechnen. Die Ergebnisse belegen jedoch, dass der Einfluss keiner der Bewertungsdimensionen der vergangenen Umfrageerfahrung auf die generalisierte Umfrageeinstellung danach divergiert, ob die Befragten ihre Angaben in einer privaten oder öffentlichen Antwortsituation gemacht haben. Keiner der Interaktionseffekte zwischen dem Befragungsmodus und den drei Bewertungsdimensionen erweist sich als signifikanter Bestimmungsfaktor der berichteten Umfrageeinstellung.

¹⁰ Dem Problem hoher Multikollinearität bei der Schätzung multiplikativer Interaktionsterme wird bei allen vorliegenden Analysen durch die Verwendung z-standardisierter und damit am Mittelwert zentrierter Variablen begegnet (Cronbach 1987). Durch diese Transformation können zwar die Schätzergebnisse von Parametern niedrigerer Ordnung beeinflusst werden, die für unser Analyseziel relevanten Parameter der jeweils höchsten Interaktionsstufe bleiben hiervon jedoch unberührt (Aiken/West 1991: 28ff.).

3.4.6 Die zeitliche Distanz der Umfrageerfahrung und die kognitive Zugänglichkeit der Erfahrungsbewertung als Moderatorvariablen

Es ließe sich vermuten, dass die Bewertungen der letzten Umfrageerfahrung mit zunehmender zeitlicher Distanz der zugrundeliegenden Erlebnisse zum Messzeitpunkt ihre Prägekraft für die generalisierte Bewertung von Umfragen verlieren.

Aus diesem Grund wird in Modell 4 überprüft, ob die Einflüsse der drei Dimensionen der Erfahrungsbewertung dann abnehmen, wenn diese Erfahrung zunehmend länger zurückliegt. Die Daten zeigen, dass dies für keine der Bewertungsdimensionen und insbesondere nicht für die wahrgenommenen Belastungen bei der letzten Befragungsteilnahme der Fall ist. Keiner der Interaktionseffekte zwischen dem zeitlichen Abstand der letzten Befragungsteilnahme und den darauf bezogenen Bewertungsdimensionen erweist sich als statistisch signifikant.

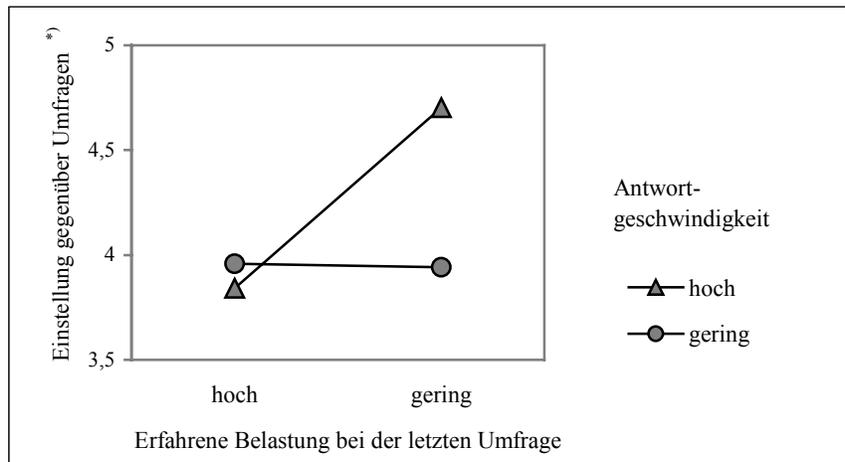
Die Ergebnisse von Regressionsmodell 4 zeigen jedoch auch, dass der kognitiven Zugänglichkeit der Erfahrungsbewertungen - wie sie in den Reaktionszeiten zum Ausdruck kommt - eine signifikante Moderatorfunktion für die Abhängigkeit der generalisierten Umfrageeinstellung von der Erfahrungsbewertung zukommt. Dies ist allerdings nur für den Interaktionseffekt zwischen den bewerteten Belastungen bei der vergangenen Umfrageteilnahme und den Antwortlatenzen bei deren Äußerung zutreffend. Die beiden anderen Bewertungsdimensionen erweisen sich selbst dann als irrelevant, wenn die Populationsheterogenität in Hinblick auf deren kognitive Verfügbarkeit berücksichtigt wird. Da in Modell 4 gleichzeitig die Bedeutung der zeitlichen Distanz statistisch kontrolliert ist, kann die Schlussfolgerung gezogen werden, dass durch die Antwortlatenzen andere Aspekte der Salienz der Umfrageerfahrung als deren gedächtnismäßige Erinnerbarkeit erfasst werden.

In Abbildung 1 ist der signifikante Interaktionseffekt zwischen der bewerteten Belastung durch die letzte Umfrageerfahrung und den bei der Äußerung dieser Urteile beobachteten Antwortgeschwindigkeiten dargestellt.¹¹ Hierbei zeigt sich, dass die generalisierte Umfrageeinstellung von Befragten mit geringer Antwortgeschwindigkeit und damit geringer kognitiver Zugänglichkeit der Erfahrungsbewertung nicht durch deren Inhalte beein-

¹¹ In Abbildung 1 wird die in Modell 4 (Tabelle 6) regressionsanalytisch vorhergesagte generalisierte Umfrageeinstellung für unterschiedliche Kombinationen der Erfahrungsbewertung und Antwortlatenzen dargestellt. Die vier Datenpunkte repräsentieren die möglichen Merkmalskombinationen von Befragten mit einer Standardabweichung über und unter der mittleren Belastungsbewertung bzw. Antwortgeschwindigkeit. Alle anderen in der Regressionsgleichung enthaltenen Effekte wurden entweder auf dem Populationsmittelwert (kontinuierliche Variablen) oder auf der jeweiligen Referenzkategorie (kategoriale Variablen) fixiert.

flusst wird. Dies ist allerdings in zunehmendem Ausmaß der Fall, wenn eine steigende Antwortgeschwindigkeit ein höheres Ausmaß an Salienz der Bewertungsinhalte anzeigt.¹²

Abbildung 1: Abhängigkeit der Einflussstärke der Belastungsbewertung auf die Umfrageeinstellung bei geringer und hoher kognitiver Zugänglichkeit



^{*)} Vorhergesagte Einstellungen gegenüber Umfragen: Skalenwert 1=extrem negative Einstellung; Skalenwert 7=extrem positive Einstellung

12 Es könnte der Einwand vorgebracht werden, dass die nach der Antwortgeschwindigkeit unterschiedliche Erklärungskraft der Erfahrungsbewertung, zumindest teilweise, ein Artefakt der Anpassung der Latenzdaten zwischen den Erhebungsmodi ist. So wurde die durchschnittliche Differenz in der Antwortgeschwindigkeit zwischen selbst- und intervieweradministrierten Interviews als Indikator für die von den Befragten benötigten Lesezeit herangezogen und diese Differenz von den Latenzen aller selbst-administrierter Interviews abgezogen. Geht man bei jüngeren und gebildeteren Befragten, im Vergleich zur jeweiligen Komplementärgruppe, von kürzeren Lesezeiten aus, so werden bei dieser Vorgehensweise deren Netto-Antwortlatenzen unterschätzt. Demnach besteht bei selbst-administrierten Interviews die Gefahr einer künstlich erzeugten Korrelation zwischen den Antwortlatenzen und sozialstruktureller Gruppenzugehörigkeit, durch die möglicherweise die Moderatorwirkung der Antwortlatenzen bewirkt wird. Zusätzlich durchgeführte Analysen bestätigen diese Alternativklärung nicht. So finden sich weder bei selbst- noch bei interviewer-administrierten Befragungen signifikante Zusammenhänge zwischen den Antwortlatenzen und der Bildung oder dem Alter der Befragten. Auch unterscheiden sich die Einflüsse der Erfahrungsbewertungen der Befragten nicht nach der Zugehörigkeit zu diesen sozialstrukturellen Gruppen (Ergebnisse nicht berichtet).

4. Zusammenfassung und Diskussion

Die vorliegende Untersuchung hat gezeigt, dass sich unterschiedliche Aspekte der Erfahrung mit Umfragen in der Vergangenheit in zum Teil deutlichem Ausmaß auf die generalisierte Einstellung der Befragten zu Umfragen auswirken. So erweist sich *erstens* die Verwendung eines persönlichen Administrationsmodus bei der Umfrage, an der die Probanden zuletzt teilgenommen haben, im Vergleich zu allen anderen Befragungsarten, als negativer Faktor für die allgemeine Bewertung von Umfragen. Dieser Effekt verschwindet jedoch dann, wenn die Bewertung dieser vergangenen Interviewerfahrung in Hinblick auf die Interviewlänge und die insgesamt aufgetretenen Anstrengungen statistisch kontrolliert wird. Es kann daher geschlussfolgert werden, dass der anfänglich beobachtete negative Einfluss einer persönlichen Art der Befragung nicht auf einer Ablehnung dieses Administrationsmodus, sondern auf der hierbei meist vorliegenden deutlich längeren Interviewdauer beruht.

Als *zweites* Analyseergebnis lässt sich feststellen, dass die Verwendung von Umfragen als „Türöffner“ für den Verkauf von Konsumprodukten keinen negativen Einfluss auf das Image von Umfragen ausübt. So zeigen Befragte, die in der Vergangenheit um ihre Teilnahme an einer Umfrage gebeten wurden, welche später in ein Verkaufsgespräch mündete, keine negativeren Einstellungen zu Umfragen als solche Probanden, die noch nie eine solche Täuschung erlebt haben. Demnach differenziert zumindest der in unserer Stichprobe repräsentierte befragungsbereite Teil der Bevölkerung nach der Verantwortlichkeit für diese negativen Erlebnisse und schreibt sie nicht einfach der Umfragebranche zu.

Die Ergebnisse der vorliegenden Studie verdeutlichen *drittens*, dass sich entgegen den bestehenden Erwartungen der Typ des Auftraggebers der letzten Befragung nicht auf die Einstellung der Befragten gegenüber Umfragen auswirkt. Die generalisierte Bewertung von Umfragen unterscheidet sich somit nicht danach, ob die Probanden zuvor an einer Befragung durch eine wissenschaftliche Einrichtung oder durch ein kommerzielles Markt- und Meinungsforschungsinstitut partizipierten. Die Hypothese, wonach sich die Legitimitätsdefizite kommerzieller Befragungen beziehungsweise der dabei vorherrschenden Befragungsthemen negativ auf die Bewertung von Umfragen auswirken könnten, lässt sich mit unseren Daten nicht bestätigen. Ein Grund hierfür kann eventuell darin gesehen werden, dass (potentielle) Befragte, die die Teilnahme an einem kommerziellen Interview tatsächlich als negative Erfahrung werten, sich erst gar nicht zur Teilnahme an einem solchen Interview bereit erklären.

Das *vierte* und wichtigste Ergebnis der Untersuchung zeigt, dass sich die drei identifizierten, voneinander unabhängigen Bewertungsdimensionen der vergangenen Interviewerfahrung in unterschiedlichem Ausmaß auf die generalisierte Einstellung zur Umfrageinstitu-

tion auswirken. Demnach lassen sich keine Einflüsse des subjektiv beurteilten Unterhaltungswertes der letzten Umfrageteilnahme oder der Wahrnehmung, inwieweit die Fragen in diesem Interview verwirrend und unklar formuliert waren, auf das generelle Image der Umfrageforschung feststellen. Allerdings finden sich relativ starke Einflüsse der erlebten Belastungen durch diese Befragungsteilnahme. Demzufolge führen aus der Sicht der Befragten zu hohe Zeitaufwendungen durch lange Fragebögen und allgemein die Wahrnehmung, dass die Befragung mit zu großen Anstrengungen verbunden war, zu einer Beeinträchtigung der Umfrageeinstellung. Dies muss insofern als wichtiges Ergebnis angesehen werden, als die bereits in Abschnitt 1 dargestellten Untersuchungsergebnisse die Schlussfolgerung nahe legen, dass sich die Einstellungen zu Umfragen - beispielsweise vermittelt über die Teilnahmebereitschaft der Befragten und das Ausmaß des damit verbundenen Non-Response Bias - auf die realisierbare Datenqualität auswirkt. Demnach kann in der Tat festgestellt werden, dass sich die Umfrageforschung zumindest teilweise, vermittelt durch das Ausmaß der Anforderungen an die Befragten, die Randbedingungen für den Erfolg ihrer zukünftigen Arbeit selbst schafft. Ein Verzicht auf lange und für die Befragten schwierige Befragungsinstrumente, oder eine zumindest symbolische Anerkennung der erbrachten Leistungen liegt somit im langfristigen Interesse der Profession.

Als *fünftes* Ergebnis verdeutlicht die Studie durch Verwendung der Antwortlatenzen bei der Bewertung der vergangenen Umfrageerfahrung, dass diese Urteile in unterschiedlichem Ausmaß kognitiv verfügbar sind und sich in Abhängigkeit von diesem Merkmal unterschiedlich stark auf die Umfrageeinstellung auswirken. So finden sich mit steigender Antwortgeschwindigkeit zunehmend stärkere Einflüsse der bewerteten Belastungen auf die Umfrageeinstellung. Auf welchen Faktoren die Unterschiede in der kognitiven Zugänglichkeit fußen, muss in der vorliegenden Untersuchung weitgehend ungeklärt bleiben. Ausgeschlossen werden kann allerdings, dass diese auf unterschiedliche zeitliche Distanzen zwischen der betreffenden Umfrageerfahrung und dem Messzeitpunkt zurückgehen. Die Bedeutung der Antwortlatenzen als Moderatorvariable für den Einfluss der Erfahrungsbewertung auf die Umfrageeinstellung bleibt auch dann erhalten, wenn gleichzeitig die mögliche Relevanz des zeitlichen Abstandes in dieser Hinsicht statistisch kontrolliert wird. Ein vermeintlicher Erklärungsfaktor für die vorliegenden Unterschiede in der kognitiven Verfügbarkeit der Bewertungsurteile könnte darin vermutet werden, dass sich besonders bei solchen Befragten die Interviewerfahrung im Gedächtnis festgesetzt hat, die sich wegen ihrer beispielsweise altersbedingten geringeren Belastbarkeit besonders stark belastet gefühlt haben. Dieser Erklärungsansatz lässt sich allerdings wegen mangelnder Stichprobengröße mit unseren Daten nicht überprüfen.

Da die vorliegende Datengrundlage in einer Zufallsstichprobe einer Großstadt besteht, lassen sich die gefundenen Randverteilungen der Umfrageeinstellung und Umfrageerfah-

rung nicht einfach auf bundesdeutsche Verhältnisse übertragen. So ist bekannt, dass in Großstädten eine Konzentration von Marktforschungsinstituten und Forschungseinrichtung vorliegt, die hier zu einer deutlich überdurchschnittlichen Umfrageerfahrung der Bevölkerung führt (McDaniel/Verille/Madden 1985: 76). Vor dem Hintergrund der somit zweifellos überdurchschnittlich vielen Erfahrung mit Umfragen in der vorliegenden Stichprobe kann daher die Frage gestellt werden, ob in Bevölkerungsteilen mit weniger intensiver Befragungstätigkeit die wahrgenommene Belastung durch Interviews das gleiche Niveau an Salienz und damit Erklärungskraft für die Einstellung zu Umfragen erreicht. Aus diesem Grund und wegen einer insgesamt besseren Absicherung der hier gefundenen Zusammenhänge durch größere Fallzahlen wäre eine Anschlussuntersuchung mit einer repräsentativen Stichprobe ein nützliches Unterfangen.

Eine andere offene Frage der vorliegenden Untersuchung betrifft die Generalisierbarkeit der gefundenen Zusammenhänge auf die Gruppe der Teilnahmeverweigerer. Da vorliegende Ergebnisse darauf hindeuten, dass gerade Befragte mit einer wirklich negativen Einstellung zu Umfragen nicht an Befragungen teilnehmen, können die hier vorgestellten Zusammenhänge nicht ohne weiteres auf diesen Teil der Grundgesamtheit verallgemeinert werden (Goyder 1986; Rogelberg/Fisher/Maynard/Hakel/Horvath 2001). Zu diesem Zweck wäre die Durchführung einer analogen Untersuchung mit einer Gruppe konvertierter Teilnahmeverweigerer und damit mit Probanden mit einer bekanntermaßen negativeren Einstellung zu Umfragen notwendig (vgl. hierzu beispielsweise Erbslöh/Koch 1988).

Es sei weiterhin angemerkt, dass die in der vorliegenden Untersuchung analysierten Bewertungsdimensionen der Umfrageerfahrung selbstverständlich keine Vollständigkeit beanspruchen. So kann beispielsweise vermutet werden, dass unterschiedliche Aspekte des Interviewerverhaltens oder die thematische Ausrichtung von Umfragen eigenständige Urteilsdimensionen und damit potentiell weitere Erklärungsfaktoren der generalisierten Bewertung von Umfragen darstellen. Da beim derzeitigen Forschungsstand für den bundesdeutschen Kontext praktisch keine systematischen Untersuchungen über die Rückkopplung der Umfragetätigkeit auf die Kooperationsbereitschaft der Informanten vorliegen, ist eine weitere empirische Vertiefung der vorliegenden Fragestellung ein lohnendes Feld zukünftiger Forschung.

Korrespondenzadressen

Bettina Langfeldt
Institut für Soziologie
Justus-Liebig-Universität Gießen
Karl-Glöckner-Str. 21E; D - 35394 Gießen
Tel.: 0641-9923242
bettina.langfeldt@sowi.uni-giessen.de

Dr. Volker Stocké
Sonderforschungsbereich 504
„Rationalitätskonzepte, Entscheidungsverhalten und ökonomische Modellierung“
TP 7 „Framing-Effekte und sozial erwünschtes Antwortverhalten“
Universität Mannheim
L13, 15; D-68131 Mannheim
Tel.: 0621-1813432
vstocke@rumms.uni-mannheim.de

Literaturverzeichnis

- Aiken, L.S./West, S.G., 1991: Multiple Regression: Testing and Interpreting Interactions. Newbury Park: Sage.
- Albaum, G., 1987: Do Source and Anonymity Affect Mail Survey Results? *Academy of Marketing Science* 15: 74-81.
- Bargh, J.A., 1996: Automaticity in Social Psychology. S.169-183 in: Higgins, E.T./Kruglanski, A.W. (Eds.): *Social Psychology. Handbook of Basic Principles*. New York: Guilford Press.
- Bassili, J.N., 1993: Response Latency versus Certainty as Indexes of the Strength of Voting Intentions in a CATI Survey. *Public Opinion Quarterly* 57: 54-61.
- Bassili, J.N., 1996: The How and Why of Response Latency Measurement in Telephone Surveys. S.319-346 in: Schwarz, N./Sudman, S. (Eds.), *Answering Questions. Methodology for Determining Cognitive and Communicative Processes in Survey Research*. San Francisco: Jossey-Bass.
- Bassili, J.N./Bors, D.A., 1997: Using Response-Latency to Increase Lead Time in Election Forecasting. *Canadian Journal of Behavioural Science* 29: 231-238.
- Bassili, J.N./Fletcher, J.F., 1991: Response-Time Measurement in Survey Research. A Method for CATI and a New Look at Nonattitudes. *Public Opinion Quarterly* 55: 331-346.
- Baxter, R., 1964: An Inquiry into the Misuse of the Survey Technique by Sales Solicitors. *Public Opinion Quarterly* 28: 124-134.
- Biel, A.L., 1967: Abuses of Survey Research Techniques: The Phony Interview. *Public*

Opinion Quarterly 31: 298.

Bradburn, N., 1979: Respondent Burden. in: Reeder, L. (Ed.): Health Survey Research Methods: Second Biennial Conference. Williamsburg, Va.: U.S. Government Printing Office.

Cronbach, L.J., 1987: Statistical Tests for Moderator Variables: Flaws in Analyses Recently Proposed. *Psychological Bulletin* 102: 414-417.

Dillman, D.A./Carpenter, E.H./Christenson, J.A./Brooks, R.M., 1974: Increasing Mail Questionnaire Response: A Four State Comparison. *American Sociological Review* 39: 744-756.

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

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

Eagly, A.H./Chaiken, S., 1993: *The Psychology of Attitudes*. Fort Worth : Harcourt Brace Jovanovich.

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

Forsa, 1991: Umfragen in Deutschland. Erfahrungen und Meinungen der Bürger 1991. *Context*, Folge 8/91: 1-14.

Forsa, 1993a: Umfragen in Deutschland. Erfahrungen und Meinungen der Bürger. Teil I. *Context*, Folge 7/93: 1-7.

Forsa, 1993b: Umfragen in Deutschland. Erfahrungen und Meinungen der Bürger. Teil II. *Context*, Folge 8/93: 1-10.

Forsa, 2000: Die Akzeptanz von Umfragen in Deutschland Anfang 2000. *Context*, Folge 7/00: 2-13.

Goyder, J., 1986: Surveys on Surveys: Limitations and Potentialities. *Public Opinion Quarterly* 50: 27-41.

Groves, R.M./Kahn, R.L. (1979): *Surveys by Telephone. A National Comparison with Personal Interviews*. New York: Academic Press

Hartfiel, G./Hillmann, K.-H., 1982: *Wörterbuch der Soziologie*. Stuttgart: Kröner.

Hartmann, E.L./Isaacson, H.L./Jurgell, C.M., 1968: Public Reaction to Public Opinion Surveying. *Public Opinion Quarterly* 32: 295-298.

Hawkins, D.I., 1979: The Impact of Sponsor Identification and Direct Disclosure of Respondent Rights on the Quantity and Quality of Mail Survey Data. *Journal of Business* 52: 577-590.

Hazzard, A., 1983: Children's Experience with, Knowledge of, and Attitude toward Dis-

abled Persons. *The Journal of Special Education* 17: 131-139.

- Jay, G.M./Willis, S.L., 1992: Influence of Direct Computer Experience on Older Adults' Attitude Toward Computers. *Journal of Gerontology: Psychological Sciences* 47: 250-257.
- Jones, C./Sheatsley, P./Stinchcombe, A., 1979: Dakota Famers and Ranchres Evaluate Crop and Livestock Surveys. Chicago: National Opinion Research Center.
- Jones, W.H./Lang, J.R., 1980: Sample Composition Bias and Response Bias in a Mail Survey: A Comparison of Inducement Methods. *Journal of Marketing Research* 18: 69-76.
- Jones, W.H./Lang, J.R., 1983: Reducing Item Omission in Mail Surveys. *Political Methodology* 9: 285-293.
- Kohut, A., 1986: Rating the Polls: The Views of Media Elites and the General Public. *Public Opinion Quarterly* 50: 1-9.
- Költringer, R., 1992: Befragungshäufigkeit und Antwortverhalten. *WISDOM* 4 (1/2): 1-13.
- Leiblein, A./Klass, J., 1990: Die Akzeptanz von Umfragen. *Planung und Analyse* 7: 259-262.
- Leiblein, A./Oglesby, S., 1993: Akzeptanz von Umfragen. Vier Länder im Vergleich. *Planung und Analyse* 20: 47-54.
- McDaniel, S.W./Verille, P./Madden, C.S., 1985: The Threats to Marketing Research: An Empirical Reappraisal. *Journal of Marketing Research* 21: 74-80.
- Nederhof, A.J., 1981: Some Sources of Artifact in Social Science Research: Nonresponse, Volunteering and Research Experience of Subjects. Leiden: Rijksuniversity Leiden.
- Nederhof, A.J., 1986: Effects of Research Experience of Respondents. *Quality & Quantity* 120: 277-284.
- Nederhof, A.J., 1987: When Neutrality is Negative: Pleasantness of Most Recent Survey experience and Nonresponse. *Quality & Quantity* 21: 425-432.
- Orr, C.J., 1976: Sponsorship, Self Presentation, and Legitimacy. *Communication Monographs* 43: 80-90.
- Peterson, R.A., 1975: An Experimental Investigation of Mail-Survey Responses. *Journal of Business Research* 3: 199-210.
- Porst, R., 1998: Erfahrung mit und Bewertung von Umfragen. Was unsere Befragten über Umfragen denken. *ZUMA-Arbeitsbericht* 3: 1-14.
- Presser, S./Blair, J./Triplett, T., 1992: Survey Sponsorship, Response Rates, and Response Effects. *Social Science Quarterly* 73: 699-702.
- Reuband, K.-H., 1998: Umfragen in Ostdeutschland. *Planung & Analyse. Zeitschrift für Informationsmanagement, Markt-, Media- und Werbeforschung* 2: 58-63.
- Rogelberg, S.G./Fisher, G.G./Maynard, D.C./Hakel, M.D./Horvath, M., 2001: Attitudes Toward Surveys: Development of a Measure and its Relationship to Respondent Behavior. *Organizational Research Methods* 4: 3-25.
- Roper, B.W., 1986: Evaluating Polls with Poll Data. *Public Opinion Quarterly* 50: 10-16.

- Rugg, W.D., 1971: Interviewer Opinion on the "Salesman as Interviewer" Problem. *Public Opinion Quarterly* 35: 625-626.
- Schleifer, S., 1986: Trends in Attitudes Toward and Participation in Survey Research. *Public Opinion Quarterly* 50: 17-26.
- Sharp, L.M./Frankel, J., 1983: Respondent Burden: A Test of Some Common Assumptions. *Public Opinion Quarterly* 47: 36-53.
- Sheets, T./Radlinski, A./Kohne, J./Brunner, G.A., 1974: Deceived Respondents: Once Bitten, Twice Shy. *Public Opinion Quarterly* 38: 261-263.
- Sjoberg, G., 1955: A Questionnaire on Questionnaires. *Public Opinion Quarterly* 18: 423-427.
- Stadt Mannheim, 1998: Elektronischer Datensatz über die Verteilung der Bevölkerungsmerkmale im Stadtgebiet, Stand 31.12.1998. Mannheim: Statistikstelle.
- Statistisches Bundesamt, 1998: Statistisches Jahrbuch für die Bundesrepublik Deutschland. Stand 31.12.1996. Wiesbaden: Online-Version: http://infob1.uni-mannheim.de/IBplus/sessions/STATJB_DEUT98-72eca49b.html.
- Statistisches Landesamt Baden-Württemberg, 1998: Erwerbstätige in Baden Württemberg 1997 nach geleisteter Wochenarbeitszeit. Baden-Württemberg in Wort und Zahl 7.
- Stinchcombe, A.L./Jones, C./Sheatsley, P., 1981: Nonresponse Bias for Attitude Questions. *Public Opinion Quarterly* 45: 359-375.
- Stocké, V., 2002: Die Vorhersage von Fragenreihenfolgeeffekten durch Antwortlatenzen: Eine Validierungsstudie. *ZUMA-Nachrichten* 50: 26-53.
- Vanberg, V.J., 2002: Rational Choice vs. Program-Based Behavior. *Alternative Theoretical Approaches and Their Relevance for the Study of Institutions. Rationality & Society* 14: 7-54.

Anhang

Tabelle 7: Die Dimensionen der Umfrageerfahrung (Ergebnisse einer Hauptkomponentenanalyse mit Varimax-Rotation)

	Rotierte Faktoren		
	1	2	3
Die Befragung hat zu lange gedauert	.210	.008	.786
Die Befragung war anstrengend	-.086	.209	.803
Die Fragen waren teilweise unklar formuliert	.065	.890	-.036
Ich habe manchmal den Sinn der Fragen nicht erkennen können	.042	.793	.297
Die Umfrage war interessant	.927	.003	-.108
Die Befragung hat Spaß gemacht	.859	.120	.283

Kommunalitäten: Faktor 1: 2.09; Faktor 2: 1.45; Faktor 3: 1.04;
Erklärte Gesamtvarianz=76.31%

BILDUNGSPARTIZIPATION UND HEIRATSNEIGUNG: DIE ENTWICKLUNG DES BILDUNGSSELEKTIVEN HEIRATSVERHALTENS IN WESTDEUTSCHLAND ZWISCHEN 1970 UND 1997

HEIKE WIRTH & SIMONE SCHMIDT

Die Höherqualifizierung von Frauen und ihre hierdurch erheblich verbesserten Berufschancen zählen zu den zentralen Faktoren des sozialen Wandels in modernen Gesellschaften. In der aktuellen gesellschaftspolitischen als auch in der - durchaus kontrovers geführten - sozialwissenschaftlichen Diskussion wird diese Entwicklung als ein Auslöser für den Wandel in Familiengründungsprozessen wie etwa dem Rückgang der Heirats- und Geburtenhäufigkeit gesehen. Der Beitrag greift die Frage auf, ob und in welchem Ausmaß die Höherqualifizierung von Frauen zu einer Veränderung der Heiratsneigung Anfang der 70er bis Ende der 90er Jahre geführt hat. Weiterhin wird geprüft, ob ein Zusammenhang zwischen dem individuellen Bildungsniveau von Männern und dem Haushaltsmodus besteht und welche Veränderungen sich hierbei im Zeitverlauf ergeben. Für das Jahr 1997 wird zusätzlich ein Ost-West-Vergleich vorgenommen.

Die vorliegenden Analysen der Volkszählungsdaten 1970 und des Mikrozensus 1997 deuten auf ein ausgeprägtes bildungsselektives Heiratsverhalten westdeutscher Frauen hin: Je höher die Qualifikation, desto geringer ist die Heiratswahrscheinlichkeit. Die Bildungsselektivität hat sich jedoch in den letzten 30 Jahren nicht vergrößert. In den alten Bundesländern sind im Vergleich zu den Frauen bei den Männern nur schwach ausgeprägte Bildungseffekte beobachtbar: Die Heiratswahrscheinlichkeit von Männern mit berufsqualifizierendem Abschluss ist höher als die derer ohne Berufsausbildung. In den neuen Bundesländern ist der Einfluss von Bildung auf das Heiratsverhalten von Frauen deutlich geringer. Die am geringsten qualifizierten Frauen haben hier die höchste Wahrscheinlichkeit, ledig zu sein. Für ostdeutsche Männer steigt dagegen die Wahrscheinlichkeit verheiratet zu sein mit dem Bildungsniveau.

This paper examines the hypothesis that the increasing educational attainment and related growing economic independence among women have led to a 'decline in marriage'. The empirical analysis focuses on marriage behaviour of women and men in West Germany from the early 1970's to the late 1990's, using the German Population Census 1970 und the German Microcensus 1997. The empirical results indicate that better educated women in general have a lower propensity to marry. However, this educational effect has not increased over time. For men the impact of education on marriage is much weaker than for women and is mainly based on vocational training: men with a vocational training are more likely to be married than men without a vocational training.

In addition we compare marriage behaviour in the eastern and western states of Germany using the German Microcensus 1997. In general the effect of educational attainment is lower for women in eastern Germany than for women in West Germany and it takes a different direction: In the eastern states women with the lowest educational attainment are most likely to be unmarried. In contrast to this, the higher the level of educational attainment, the more likely it is that males in the eastern states are married.

1. Einleitung

Ausgehend von der familienökonomischen Theorie von Gary S. Becker wird in der Literatur seit längerem über die Konsequenzen der erhöhten Bildungspartizipation von Frauen und der damit einhergehenden stärkeren beruflichen Orientierung auf das Heiratsverhalten diskutiert.¹ Im Zentrum des Interesses steht hierbei insbesondere die These Beckers (1981: 248), dass "*the gain from marriage is reduced by a rise in the earnings and labor force participation of women (...), because a sexual division of labor becomes less advantageous.*"

Diese These wird im Allgemeinen dahingehend interpretiert, dass ein solcher Effekt vor allem für hochqualifizierte Frauen zu erwarten ist, da diese - aufgrund ihres höheren Einkommenspotenzials - unter der Randbedingung einer traditionellen geschlechtsspezifischen Arbeitsteilung bei einer Eheschließung mit höheren Opportunitätskosten kalkulieren müssen als weniger qualifizierte Frauen. Empirisch sollte sich dies darin zeigen, dass hochqualifizierte Frauen eine geringere Heiratswahrscheinlichkeit aufweisen, d.h. der Anteil der Ledigenquoten mit zunehmenden Ausbildungsniveau steigt. Die hierzu bislang vorliegenden empirischen Befunde ergeben allerdings ein eher widersprüchliches Bild. So finden sich in einer Reihe von Studien Belege für einen Bildungseffekt (z.B. Brüderl/Klein 1993; Brüderl/Diekmann 1994). Anderen Untersuchungen zufolge führt ein

1 Ein Überblick zur internationalen Diskussion findet sich bei Oppenheim Mason/Jensen (1995).

höheres Bildungsniveau aufgrund der längeren Verweildauer in Bildungsinstitutionen zwar zu einer Verzögerung des Heiratsprozesses, nicht jedoch zu einer Reduzierung der Heiratswahrscheinlichkeit (Oppenheimer 1988, 1995, 1997, Oppenheimer et al. 1995, Oppenheimer/Lew 1995; Blossfeld/Jaenichen 1990; Blossfeld et al. 1993; Huinink 1989).

Angesichts der aktuellen Diskussion über eine sich in Deutschland abzeichnende Bipolarisierung der Bevölkerung in einen familialen und nichtfamilialen² Sektor (Dorbritz/Gärtner 1995: 352; Ostner 1999) gewinnt die These Beckers an gesellschaftspolitischer Brisanz. Sollte es zutreffen, dass insbesondere hochqualifizierte Frauen vermehrt zugunsten der eigenen Berufstätigkeit auf eine Ehe (und damit typischerweise auch auf Kinder)³ verzichten, würde dies, da auch die Paarbildung in einem engen Zusammenhang mit dem Bildungsniveau steht (Ziegler 1985; Wirth 1996, 2000; Blossfeld/Timm 1997), langfristig mit einer bildungsselektiven Differenzierung des familialen und nichtfamilialen Sektors einhergehen. Die im Zusammenhang mit der 'Familienarbeit' direkt anfallenden Kosten wären dann nicht zufällig über alle sozio-ökonomischen Bevölkerungsgruppen verteilt, sondern wären überproportional von den sozial weniger privilegierten Gruppen zu tragen (Strohmeier 1993; Burkart 1997: 290).

Im vorliegenden Beitrag greifen wir die Frage nach dem Zusammenhang zwischen Bildungsniveau und Heiratswahrscheinlichkeit von Frauen daher erneut auf. Im Zentrum des Interesses steht die Frage, ob ein Bildungseffekt nachweisbar ist und falls ja, ob sich dieser Effekt im Verlauf der Bildungsexpansion verstärkt. Zu diesem Zweck untersuchen wir die bildungsspezifischen Heiratswahrscheinlichkeiten der 25- bis 44-jährigen Frauen Anfang der 1970er und Ende der 1990er Jahre in Westdeutschland. In Hinblick auf das Untersuchungsinteresse ist hierbei vor allem wichtig, dass die Ausbildungsphase der 1970 25- bis 44-Jährigen vor der Bildungsexpansion abgeschlossen war und in einer Zeit lag, die noch sehr stark durch traditionelle Rollenbilder und geschlechtsspezifische Normen geprägt war. Im Unterschied hierzu fiel die Ausbildung der Ende der 1990er Jahre 25- bis 44-Jährigen mit der Bildungsexpansion zusammen. Diese Epoche war nicht nur durch einen überproportionalen Anstieg der weiblichen Bildungspartizipation gekennzeichnet, sondern ging auch mit einem Abbau geschlechtsspezifischer Ungleichheitsstrukturen in anderen Lebensbereichen sowie einer erheblichen Modernisierung der 'Frauenrolle' einher. Der Zeitvergleich bietet daher sowohl die Möglichkeit, die neuere Entwicklung des

² Hierzu zählen Singles und kinderlose nichteheliche Lebensgemeinschaften, aber auch kinderlose Ehen.

³ Ungeachtet der zunehmenden Zahl von Alleinerziehenden ist die Realisierung eines Kinderwunschs in Deutschland auch in der Gegenwart in der überwiegenden Mehrheit an die Ehe gekoppelt.

Heiratsverhaltens als auch die Effekte der Bildungsexpansion zu untersuchen. Die sich daran unmittelbar anschliessende Frage der bildungsselektiven Familiengründung kann in diesem Beitrag allerdings nicht weiter verfolgt werden. Stattdessen wird die Analyse der bildungsselektiven Heiratswahrscheinlichkeit zum einen durch die Einbeziehung von Männern,⁴ zum anderen durch einen Ost-West-Vergleich erweitert. Unter theoretischen Gesichtspunkten ist dabei insbesondere der Ost-West-Vergleich von Interesse, da die institutionellen und familienpolitischen Rahmenbedingungen in der ehemaligen DDR erheblich günstigere Voraussetzungen für die Berufstätigkeit von Frauen boten als die alte Bundesrepublik.

Im Folgenden werden zunächst die theoretischen Überlegungen skizziert, dann erfolgt eine Beschreibung der Datenbasis und der verwendeten Merkmale. Im dritten Abschnitt werden die Analysemodelle und die Ergebnisse präsentiert. Abschliessend werden die zentralen Befunde kurz zusammengefasst.

2. Geschlechtsspezifische Arbeitsteilung und Heiratsneigung: Der mikroökonomische Erklärungsansatz

Bei den folgenden Ausführungen konzentrieren wir uns auf den familienökonomischen Erklärungsansatz von Gary S. Becker (1981, 1991). Wir sind uns dabei bewusst, dass andere soziologische und - Becker zum Teil modifizierende - ökonomische Ansätze⁵ zur Erklärung der geschlechtsspezifischen Arbeitsteilung vorliegen, wie auch dessen, dass Beckers Erklärungsmodell nicht unumstritten ist (Ben-Porath 1982; Hannan 1982; Oppenheimer 1988, 1997; Oppenheim Mason/Jensen 1995; Pies/Leschke 1998). Aber trotz mancher berechtigter Kritikpunkte ist auch die soziologische Diskussion über die Mechanismen des empirisch zu beobachtenden Wandels im Heirats- und Familiengründungsverhalten, stark durch Beckers ökonomische Theorie der Familie beeinflusst. Da die Grundzüge der familienökonomischen Theorie als bekannt vorausgesetzt werden können, werden im Folgenden nur knapp die zentralen Aussagen zur geschlechtsspezifischen Arbeitsteilung als Basis für die Erklärung von Heiratsverhalten skizziert.

In einfachen Worten geht der familienökonomische Ansatz davon aus, dass Personen dann eine Ehe eingehen, wenn beide Partner hierdurch im Vergleich zu getrennt wirtschaftenden Singlehaushalten einen höheren Nutzen erzielen. Nach Becker beruht der materielle

⁴ Von Ausnahmen (z.B. Huinink 1995; Brüderl/Diekmann 1994) abgesehen, konzentrieren sich die vorliegenden Studien zum Heiratsverhalten in Deutschland auf Frauen. Das Verhalten von Männern als eine Erklärung für den Wandel in Lebensformen wird eher selten berücksichtigt (Tölke 1995).

⁵ Siehe u.a. Ott (1998); Röhler et al. (2000); Seel (2001).

Vorteil eines (Ehe-)Paarhaushaltes dabei im Wesentlichen auf sogenannten Spezialisierungsgewinnen, die aus einer arbeitsteiligen Organisation von Haus- und Erwerbsarbeit resultieren. Zentral hierfür ist, dass in einem Paarhaushalt im Unterschied zu einem Singlehaushalt Partnerressourcen gepoolt werden können. Dies erlaubt eine größtmögliche Spezialisierung der Partner auf einen Zuständigkeitsbereich, d.h. entweder auf den Arbeitsmarkt oder auf Hausarbeit, wodurch sich die individuelle 'Produktivität' in dem jeweiligen Bereich erhöht. Gegeben diese Spezialisierung, liegt der durch eine gemeinsame Haushaltsführung zu erzielende Gewinn über dem von zwei Singlehaushalten in der Summe realisierbaren Gewinn.

Die Erweiterung des Erklärungsmodells um die Annahme, dass der wesentliche Zweck einer Ehe darin besteht, Kinder zu haben und aufzuziehen (Becker 1991: 135), ermöglicht dann die Folgerung, dass die *traditionelle geschlechtsspezifische Arbeitsteilung* - gemessen an ökonomischen Kriterien - am effizientesten ist. Denn unabhängig von geschlechtsspezifischen Normen oder Rollenzuschreibungen kann unter dieser Randbedingung für Frauen - aufgrund ihrer Gebärfähigkeit - eine biologisch bedingte, wenn auch zunächst möglicherweise nur geringfügig höhere relative Produktivität im Hausarbeitsbereich angenommen werden (Becker 1991: 37f.). Dem Modell zufolge ist es daher für Paarhaushalte naheliegend, dass Frauen diesen komparativen Vorteil durch verstärkte Investitionen in haushaltsspezifisches Humankapital ausbauen, während Männer vorwiegend in arbeitsmarktrelevantes Humankapital, d.h. in ihre Bildung und Ausbildung investieren, um ihre Arbeitsmarktproduktivität zu erhöhen. Letzteres führt wiederum dazu, dass Männer im Vergleich zu Frauen einen komparativen Vorteil am Arbeitsmarkt haben, d.h. ein höheres Einkommen erzielen. Unter diesen Voraussetzungen besteht der materielle Hauptanreiz für eine Eheschließung demnach in einer spezifischen Austauschbeziehung zwischen Männern und Frauen, welche dadurch gekennzeichnet ist, dass sich die Frau auf die Haushaltsführung und Kindererziehung spezialisiert, während der Mann das hierfür notwendige Einkommen erwirtschaftet. Ausgehend von diesen Überlegungen formuliert Becker (1981: 248) im Weiteren die bereits eingangs erwähnte prominente These, nach welcher höhere Einkommen und eine höhere Erwerbsbeteiligung von Frauen die Vorteile der geschlechtsspezifischen Spezialisierung reduzieren und dadurch den Anreiz zu einer Eheschließung verringern.

Diese These wurde in weiten Teilen der familiensoziologischen Forschung mit Interesse aufgenommen. Ermöglicht sie doch, zwei in den letzten 30 Jahren in vielen modernen Gesellschaften zu beobachtende soziodemographische Trends, nämlich (1.) die rapid wachsende Bildungs- und Erwerbsbeteiligung von Frauen und (2.) die sinkenden Heirats- und Geburtenziffern, theoretisch begründet in einen kausalen Zusammenhang zu stellen. In der Folge wurde eine Reihe von empirischen Untersuchungen zu dieser Thematik

veröffentlicht. Diese stimmen in der Tendenz dahingehend überein, dass höhere Bildung aufgrund der längeren Verweildauer in Bildungsinstitutionen mit einer verzögerten Eheschließung einhergeht, d.h. das Heiratsalter mit zunehmendem Bildungsniveau steigt. Man spricht hier auch von dem sogenannten *Institutioneneffekt*. Im Unterschied hierzu hat die Frage, ob der von Becker postulierte *Humankapital- bzw. Bildungseffekt*⁶ einer empirischen Überprüfung standhält, eine lebhafte Diskussion ausgelöst. So finden sich Untersuchungen, deren Befunde gegen die These eines kausalen Zusammenhangs zwischen der Höherqualifizierung von Frauen und der rückläufigen Heiratsneigung sprechen (z.B. Blossfeld/Huinink 1989; Blossfeld/Jaenichen 1990; Oppenheimer 1997). Dem widersprechen jedoch andere Studien, denen zufolge sehr wohl ein negativer Humankapitaleffekt, d.h. eine reduzierte Heiratsneigung von höherqualifizierten Frauen, nachweisbar ist (z.B. Brüderl/Klein 1993; Brüderl/Diekmann 1994). Ohne auf diese Diskussion im Detail weiter einzugehen, können diese scheinbar widersprüchlichen Ergebnisse möglicherweise auch durch (auf den ersten Blick nicht ersichtliche) Nuancen der jeweils konkreten Forschungsfrage bedingt sein. Denn unseres Erachtens macht es durchaus einen Unterschied, ob sich das Untersuchungsinteresse auf die Überprüfung der These eines im Zeitverlauf *zunehmenden* Humankapitaleffekts richtet oder ob eher die *generelle Existenz eines Humankapitaleffekts* überprüft werden soll. Darüber hinaus erfolgt die allgemeine Diskussion von Beckers Thesen oftmals vor dem Hintergrund der allgemein verstärkten Bildungsbeteiligung von Frauen in den letzten 30 Jahren, das eigentliche Untersuchungsinteresse konzentriert sich dann jedoch auf die Heiratsneigung von hochqualifizierten Frauen. Vor diesem Hintergrund versuchen wir im Folgenden, die uns interessierenden Forschungshypothesen näher zu präzisieren.

Folgt man der familienökonomischen Argumentation, wäre erstens zu erwarten, dass die Heiratswahrscheinlichkeit von Frauen in einem negativen Zusammenhang mit der Höhe ihres Ausbildungsniveaus (als Indikator für arbeitsmarktrelevantes Humankapital) steht (**These 1**). Dies lässt sich damit begründen, dass höher qualifizierte Frauen über bessere Berufs- und Einkommenschancen verfügen und deshalb aufgrund ihrer höheren Opportunitätskosten von einer traditionellen geschlechtsspezifischen Arbeitsteilung weniger profitieren als Frauen mit einer geringeren Qualifikation (Oppenheimer/Lew 1995: 130). Dabei teilen wir allerdings nicht die Auffassung, nach welcher dieser Humankapitaleffekt gewissermaßen erst als eine Folge der allgemeinen Höherqualifizierung der Frauen zu beobachten sein sollte. Wie aus älteren Studien zur Erwerbsbeteiligung von Frauen hervorgeht, bestand auch schon Anfang der 1970er Jahre, also noch bevor die

⁶ Als Indikator für das Einkommenspotenzial wird im Allgemeinen das arbeitsmarktrelevante Humankapital, d.h. das Bildungs- und Ausbildungsniveau herangezogen.

Bildungsexpansion in vollem Umfang einsetzte, ein positiver Zusammenhang zwischen Qualifikationsniveau und Erwerbstätigkeit: die höchste Erwerbsbeteiligung findet sich bei Frauen mit einem akademischen oder mittleren berufsqualifizierenden Abschluss (Handl 1978: 229). Sofern die Spezialisierungsthese von Becker zutrifft, müsste sich das Problem der höheren Opportunitätskosten bei einer Eheschließung und Familiengründung demnach auch schon in den 70er Jahren gestellt haben und sich in Form einer mit zunehmender Qualifikation ansteigenden Ledigenquote äussern. Weiterhin ist dann zu fragen, weshalb sich die Opportunitätskosten für hochqualifizierte Frauen relativ zu weniger qualifizierten Frauen in den letzten 30 Jahren erhöht haben sollten, d.h. warum der Humankapitaleffekt im Zeitverlauf zunehmen sollte. Vielmehr sind eine Reihe von gesellschaftlichen Modernisierungsprozessen zu beobachten, welche die komparativen Vorteile der Hausfrauenehe in Hinblick auf die materielle Versorgung ganz allgemein reduzieren (Ott 1998). Hierzu gehört u.a., dass eine Vielzahl von ursprünglich im Haushalt produzierten Gütern zwischenzeitlich kostengünstig durch entsprechende Marktgüter und Dienstleistungen (z.B. in Form von Fertiggerichten, Haushaltshilfen oder Tagesmüttern) ersetzt werden können, ebenso wie der vermehrte Einsatz von komfortablen Haushaltsgeräten zu einer erheblichen Vereinfachung der Hausarbeit beigetragen hat.⁷ Hinzu kommt, dass die Einkommensunterschiede zwischen Männern und Frauen in den letzten Jahrzehnten erheblich zurückgegangen sind,⁸ wobei hochqualifizierte Frauen von dieser Entwicklung weniger profitieren als geringer qualifizierte Frauen (Steiner/Lauer 2000). Vor diesem Hintergrund ist eine Spezialisierung von Frauen auf Hausarbeit im Sinne der familienökonomischen Theorie generell immer weniger notwendig bzw. nur während der Kindererziehungsphasen rentabel. In der Tendenz erwarten wir daher *keinen zunehmenden Humankapitaleffekt*, sondern dieser gesellschaftliche Wandel sollte sich vielmehr in einem *Anstieg der Ledigenquoten über alle Bildungsgruppen* hinweg äussern (**These 2**).

⁷ Bemerkenswerterweise hat die Technisierung der Hausarbeit, d.h. der umfassende Einsatz von zeit- und arbeitssparenden Geräten, primär zu einer Flexibilisierung der Arbeitsstrukturen im Haushalt geführt, wodurch sich die Doppelbelastung durch Haus- und Erwerbsarbeit zeitlich besser managen lässt. Die für die Erledigung von Hausarbeit insgesamt aufgewandte Zeit ist dagegen allenfalls geringfügig zurückgegangen, da das Anspruchsniveau gestiegen ist und die durch die Haushaltsgeräte eingesparte Zeit nun im Wesentlichen in eine Qualitätserhöhung der Hausarbeit einfließt (Dörr/Glatzer 1995: 525).

⁸ Nach einer Pressemitteilung des Statistischen Bundesamts (6. März 2002) erzielen Arbeiterinnen im Jahr 2001 knapp 74% (1957: 57%), weibliche Angestellte knapp 71% (1957: 55%) des jeweiligen männlichen Erwerbseinkommens.

Die bisherige Argumentation konzentriert sich auf die bildungsselektive Heiratsneigung von Frauen. Wenngleich in der Literatur eher selten thematisiert, ist im Weiteren kurz zu fragen, welche Thesen die Familienökonomie in Hinblick auf das Heiratsverhalten von Männern nahe legt. Im Umkehrschluss zu der obigen Argumentation ist gemäß der familienökonomischen Theorie zu erwarten, dass die aus einer geschlechtsspezifischen Arbeitsteilung resultierenden komparativen Vorteile für Männer mit zunehmenden Einkommenspotenzial (bzw. Qualifikationsniveau) ansteigen (Brüderl/Diekmann 1994; Huinink 1995). Mit anderen Worten, während für Frauen oben ein negativer Humankapitaleffekt postuliert wurde, sollte sich für Männer ein positiver Humankapitaleffekt zeigen, d.h. die Heiratswahrscheinlichkeit sollte mit steigendem Qualifikationsniveau zunehmen (**These 3**). In Hinblick auf die Entwicklung im Zeitverlauf ist davon auszugehen, dass die oben angedeuteten Modernisierungsprozesse nicht nur bei Frauen, sondern auch bei Männern den Anreiz für eine Eheschließung über alle Bildungsgruppen hinweg verringern (**These 4**), da es sich hier um wechselseitige Prozesse handelt: Denn auch für Männer reduzieren sich - dem familienökonomischen Modell zufolge - die aus einer Ehe resultierenden Gewinne, wenn die Spezialisierung von Frauen auf die Hausarbeit und Kindererziehung abnimmt, zumal die affektiven Aspekte einer Partnerschaft in der Gegenwartsgesellschaft auch ohne eine formale Eheschließung realisierbar sind.

Bevor diese Thesen einer empirischen Überprüfung unterzogen werden, wollen wir im Folgenden noch kurz zwei Aspekte ansprechen, die bei der bisherigen Diskussion vernachlässigt wurden.

Wie eingangs dargestellt richtet sich das Untersuchungsinteresse der vorliegenden Arbeit auf die Entwicklung der bildungsselektiven Heiratsneigung im Zeitverlauf und nicht darauf, ob es sich bei der Ehe um ein gesellschaftliches Auslaufmodell handelt. Dennoch soll an dieser Stelle darauf hingewiesen werden, dass das in Beckers familienökonomischem Ansatz als optimal dargestellte Modell der traditionellen Hausfrauenehe in den meisten modernen Gesellschaften sicherlich ein Auslaufmodell ist. Dies ist allerdings nicht gleichbedeutend damit, dass die Ehe als gesellschaftliche Institution am Verschwinden ist (vgl. z.B. Nave-Herz 1988). Vielmehr ist die höhere ökonomische Unabhängigkeit von Frauen infolge ihrer verstärkten Erwerbsbeteiligung nicht nur von steigenden Ledigenquoten begleitet, sondern parallel hierzu hat auch ein Bedeutungswandel in der Funktion der Ehe stattgefunden: Weg von der Versorger- bzw. Hausfrauenehe hin zu einer eher gleichberechtigten Partnerschaft, bei welcher die Befriedigung emotionaler, physischer und psychischer Bedürfnisse im Vordergrund steht und sich beide Partner an Haus-

und Erwerbsarbeit - wenn auch selten gleichmässig - beteiligen.⁹ Insbesondere der letztere Aspekt einer gemeinschaftlichen Erledigung von Haus- und Erwerbsarbeit lässt sich mit Beckers Modell allerdings nicht erklären (Ott 1998: 72; Oppenheim Mason/Jensen 1995: 5).¹⁰

Abschliessend sei noch ein weiterer wichtiger Aspekt angesprochen. Der familienökonomischen Theorie zufolge sind die Heirats- und Familiengründungsprozesse das Ergebnis von individuellen Entscheidungsprozessen. Kulturelle, normative und institutionelle Rahmenbedingungen als zusätzliche, das individuelle Verhalten beeinflussende Faktoren sind in dem Erklärungsmodell nicht vorgesehen.¹¹ Diese Annahme erscheint aus unserer Perspektive insofern realitätsfern, als sich aus solchen Rahmenbedingungen sowohl Restriktionen wie auch Optionen für individuelle Entscheidungsprozesse ergeben können, die unter sozialpolitischen Gesichtspunkten durchaus von Relevanz sind. Um die Bedeutung derartiger Einflussfaktoren auf das bildungselektive Heiratsverhalten zumindest grob abzuschätzen, bietet sich ein Vergleich zwischen Ost- und Westdeutschland an. Dies insbesondere deshalb, als in Ostdeutschland bis zur Wiedervereinigung mittels umfangreicher sozialpolitischer Massnahmen (Wohnungsvergabe und zinsverbilligte Kredite an Ehepaare, umfangreiche Kinderbetreuungsmöglichkeiten, Erziehungsurlaub etc.) die Eheschließung und Familiengründung unterstützt wurde, bzw. die Opportunitätskosten für erwerbstätige Frauen gesenkt wurden (Brüderl/Klein 1994: 60).

3. Datenbasis und Variablen

Die folgenden Auswertungen basieren auf Daten der Volkszählung 1970 (VZ70) und des Mikrozensus¹² 1997 (MZ97). Die Daten der VZ70 sind eine Ein-Prozent-Substichprobe der Volks- und Berufszählung 1970 (Schimpl-Neimanns/Frenzel 1995); die Daten des MZ97 sind eine 70 Prozent Substichprobe des Mikrozensus 1997 (Lechert et al. 2000). Aus

⁹ Diese Bedürfnisse lassen sich natürlich gleichermassen in nicht-ehelichen Lebensgemeinschaften realisieren. Sofern jedoch der Wunsch nach einer (hier nicht diskutierten) Familiengründung hinzukommt, werden zumindest in Deutschland nach wie vor die von der Rechtsinstitution 'Ehe' ausgehenden Sicherheiten präferiert, d.h. es wird überwiegend geheiratet. Mit anderen Worten: Die Ehe "wird als Institution dann benutzt, wenn es mit der Geburt von Kindern darum geht, das höchste Maß an Sicherheit für Frauen und Kinder zu erreichen" (Höhn/Dorbritz 1995:150).

¹⁰ Seel (2001) schlägt vor, Beckers Modell so zu modifizieren, dass es mit den empirischen Befunden in Einklang gebracht werden kann.

¹¹ Vgl. z.B. Hannan (1982: 70f.).

¹² Der Mikrozensus ist eine jährlich stattfindende repräsentative Erhebung des Statistischen Bundesamtes, in der ein Prozent der Bevölkerung nach Merkmalen zum Arbeitsmarkt und zur sozio-ökonomischen Struktur befragt wird (Lüttinger/Riede 1997). Für Beispiele zum Analysepotenzial des Mikrozensus siehe Schimpl-Neimanns (1998; 2002).

bekanntem Gründen enthält die VZ70 nur Informationen über die Bevölkerung für das Gebiet der alten Bundesrepublik. Für das Gebiet der neuen Bundesländer stehen vergleichbare Datenquellen für diesen Zeitraum (z.B. die DDR-Volkszählungen) bislang nicht in Form von Scientific Use Files zur Verfügung. Deshalb konzentriert sich der Zeitvergleich zwischen 1970 und 1997 auf Westdeutschland. In einem zweiten Schritt wird dann ein Vergleich zwischen Ost- und Westdeutschland auf Basis des MZ97 vorgenommen. Weiterhin beschränken sich die Analysen auf die deutsche Bevölkerung, da für das Heiratsverhalten der ausländischen Bevölkerung andere Randbedingungen anzunehmen sind. Der aufgrund der großen Fallzahlen von VZ70 und MZ97 kleine Stichprobenfehler ermöglicht die Untersuchung von Veränderungen im Heiratsverhalten auch kleiner Bildungsgruppen (z.B. Abiturienten ohne Berufsausbildung), die bei der Verwendung von sozialwissenschaftlichen Erhebungen häufig aufgrund der geringen Zellbesetzungen mit anderen Gruppen zusammengefasst werden müssen.

Als Indikator für die Heiratsneigung wird der Haushaltsmodus der Befragten herangezogen, wobei zwischen zwei Ausprägungen unterschieden wird: verheiratet zusammenlebend versus ledig. Da die Gruppe der Ledigen in anderen Untersuchungen zum bildungsselektiven Heiratsverhalten nicht weiter danach differenziert wird, ob es sich um einen Singlehaushalt oder eine nichteheliche Lebensgemeinschaft handelt, wird auf diese Unterscheidung hier gleichfalls verzichtet, um eine annähernde Vergleichbarkeit der Befunde zu erreichen. Darüber hinaus gab es 1970 noch nahezu keine 'offiziellen' nichtehelichen Lebensgemeinschaften.¹³ Sofern unverheiratete Paare in einem Haushalt zusammen lebten, wurden sie von der amtlichen Statistik nicht als Partner erfasst. Eine Differenzierung zwischen nichtehelichen Lebensgemeinschaften und sonstigen gemeinsam in einem Haushalt lebenden ledigen Männern und Frauen wäre deshalb mit den Daten der VZ70 nur sehr grob möglich. Für 1997 zeigen (hier nicht dokumentierte) Modellüberprüfungen, dass sich Singles und nichteheliche Lebensgemeinschaften im Hinblick auf die hier interessierenden Merkmale nur hinsichtlich des Alters unterscheiden: Personen in nichtehelichen Lebensgemeinschaften sind im Schnitt etwas jünger als Singles.

Als Indikator für das individuelle Humankapital werden die Angaben zum höchsten erreichten allgemeinbildenden und berufsbildenden Abschluss in Anlehnung an die CASMIN-Klassifikation (Brauns/Steinmann 1999)¹⁴ kombiniert. Im Einzelnen

¹³ Bis Mitte der siebziger Jahre machten sich Vermieter strafbar, wenn sie unverheirateten Paaren eine Wohnung vermieteten. Der sogenannte 'Kuppelparagraph' wurde erst 1974 abgeschafft.

¹⁴ Bei der hier verwendeten Klassifikation handelt es sich um eine aktualisierte Form der Bildungsklassifikation, die im Rahmen des CASMIN-Projektes entwickelt wurde (König et al. 1987; Braun/Müller 1997).

unterscheiden wir zwischen sieben Bildungskategorien: Hauptschulabschluss ohne Berufsausbildung (*HSO*) und mit Berufsausbildung (*HSM*); Mittlere Reife ohne Berufsausbildung (*MRO*) und mit Berufsausbildung (*MRM*); Abitur ohne Berufsausbildung (*ABIO*) und mit Berufsausbildung (*ABIM*) sowie Fachhochschul- oder Hochschulabschluss (*FH/HS*).¹⁵ Personen, die sich zum Erhebungszeitpunkt noch in einer (allgemeinbildenden oder beruflichen) Ausbildung befanden, werden aus der Analyse ausgeschlossen, da - wie oben ausgeführt - während der Ausbildungsphase normative und vor allem ökonomische Randbedingungen typischerweise einer Eheschließung und Familiengründung entgegenwirken (Thornton et al. 1995).

Die Analysen konzentrieren sich auf die Gruppe der 25- bis 44-Jährigen zum Erhebungszeitpunkt 1970 bzw. 1997, wobei innerhalb dieser Gruppen nochmals eine Feindifferenzierung vorgenommen wird (vgl. Übersicht 1):

Übersicht 1: Alters- und Geburtskohortenabgrenzung für die Analyse der bildungsspezifischen Heiratsneigung

Alterskohorten	Erhebungszeitpunkt:	
	1970	1997
	Geburtskohorten	
25-29	1941 - 1945	1968 - 1972
30-34	1936 - 1940	1963 - 1967
35-39	1931 - 1935	1958 - 1962
40-44	1926 - 1930	1953 - 1957

Die Wahl dieser Altersabgrenzung bietet verschiedene Vorteile: Im Alter zwischen 25 bis 44¹⁶ Jahren ist die Ausbildungsphase meist abgeschlossen, ebenso wie ein Großteil der Eheschließungen. Über den Vergleich von Alterskohorten zu zwei verschiedenen Erhebungszeitpunkten können zudem eventuell bestehende Alterseffekte kontrolliert werden. In Bezug auf die Fragestellung ist allerdings besonders wichtig, dass die Kohorte der 1926 bis 1945 Geborenen ihre Ausbildung bis Ende der 1960er Jahre, d.h. vor der Bildungsexpansion beendet hat und in die Familiengründungsphase eingetreten ist, also in

¹⁵ Diese Zusammenfassung war aufgrund der geringen Fallzahl von Fachhochschulabsolventinnen für den Erhebungszeitpunkt 1970 notwendig, die bei einer weiteren Aufspaltung z.B. nach Alterskohorten sehr schnell zu nicht mehr aussagekräftigen Zellbesetzungen führt.

¹⁶ Diese weite Altersspanne wurde gewählt, weil in den vergangenen dreißig Jahren die Verweildauer in Ausbildungsinstitutionen stark zugenommen hat und sich parallel hierzu das Heiratsalter insgesamt erhöht hat (Huinink 1989: 135; Blossfeld 1995: 9; Brüderl/Klein 1993: 211).

einer Zeit, die noch stark von der traditionellen Frauenrolle geprägt war. Im Unterschied hierzu hat die Kohorte der 1953 bis 1972 Geborenen ihre Ausbildung zwischen den 70er und 90er Jahren abgeschlossen und in vollem Maße von der Bildungsexpansion und den gesellschaftlichen Modernisierungsprozessen, insbesondere in Hinblick auf die Einbeziehung von Frauen in die Erwerbsarbeit, profitiert. In Hinblick auf die ostdeutschen Alterskohorten ist davon auszugehen, dass für die zwei älteren - hier betrachteten - Geburtskohorten (1953 - 1962) die Ausbildungs- und Familiengründungsphase im Wesentlichen unter den Randbedingungen der ehemaligen DDR erfolgte. Sofern also gesellschaftspolitische Randbedingungen Einfluss auf das Heiratsverhalten zeigen sollten, müsste sich dies in entsprechenden Unterschieden zwischen den Ost- und Westkohorten zeigen. Ein Überblick über die Randverteilungen der zentralen Merkmale findet sich in Tabelle A1 im Anhang.

4. Ergebnisse

4.1 Die bildungsselektive Heiratsneigung 1970 und 1997 (Westdeutschland)

Zur Überprüfung der ersten Frage nach der Entwicklung der bildungsselektiven Heiratsneigung werden verschiedene binäre logistische Regressionen durchgeführt, die es ermöglichen, den jeweiligen Einfluss der hier spezifisch interessierenden Variablen 'Bildungsniveau', 'Geschlecht', 'Altersgruppe' und 'Beobachtungszeitpunkt' auf die Heiratsneigung zu überprüfen. Mit dem binären Logit-Modell wird dabei die Wahrscheinlichkeit einer Person, einem bestimmten Haushaltsmodus anzugehören ($Y=j$), in Abhängigkeit von den erklärenden Variablen (x_i) berechnet:

$$P(Y = j) = \frac{e^{x_i}}{1 + e^{x_i}} \quad \text{für } x_i \text{ gilt } \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_K x_K$$

j = ledig, verheiratet; i = 1, ..., K

Über den Vergleich der Einzelmodelle kann der relative Erklärungsbeitrag der verschiedenen Variablen auf das Heiratsverhalten und die Veränderungen über die Zeit bestimmt werden (vgl. z.B. Andreß et al. 1997: 287ff.). In Übersicht 2 sind die Modellanpassung und die Devianzwerte dokumentiert. Die Spalten der Modellanpassung zeigen für jedes Modell die Devianz, gemessen als Log-likelihood G^2 , die Zahl der Freiheitsgrade (DF), das Signifikanzniveau (P) und die erklärte Devianz der Individualdaten (Pseudo- R^2). Die Spalten der Devianzaufklärung geben an, wie sich die Varianz verändert, wenn einzelne Variablen aus dem Basismodell ausgeschlossen werden, d.h. hier werden die Nettoeffekte der vier Variablen auf den Haushaltsmodus sichtbar.

In einem ersten Schritt (Teiltabelle A) werden zunächst die Haupteffekte betrachtet. Das Unabhängigkeitsmodell (A0) postuliert, dass das Heiratsverhalten unabhängig vom Bildungsniveau, der Altersgruppenzugehörigkeit, dem Geschlecht und dem Beobachtungszeitpunkt ist. Das Ergebnis des Modells ($G^2 = 40114$; $DF = 111$) zeigt, dass die Annahme der statistischen Unabhängigkeit verworfen werden kann. Im Basismodell (A1) wird die Unabhängigkeitsthese zurückgewiesen und stattdessen unterstellt, dass die Heiratsneigung in Abhängigkeit vom Bildungsniveau, der Alterskohorte, dem Geschlecht und dem Beobachtungszeitpunkt variiert. Die stark reduzierte Devianz ($G^2 = 5028$; $DF = 100$) zeigt einen deutlichen Einfluss dieser Faktoren auf den Haushaltsmodus. Von empirisch maximal 16,4 % Gesamterklärungskraft des saturierten Modells werden durch das Basismodell bereits 14,4 % aufgeklärt. Mit den Modellen A2 bis A5 wird die partielle Erklärungskraft der einzelnen unabhängigen Variablen erfasst. Wie aus den Devianzwerten ersichtlich, tragen alle vier Variablen signifikant zur Erklärung des Haushaltsmodus bei, variieren allerdings in ihrer Erklärungskraft. Die höchste Erklärungsleistung findet sich für das Merkmal Altersgruppenzugehörigkeit, gefolgt vom Beobachtungszeitpunkt, dem Geschlecht und schliesslich dem Bildungsniveau. Mit anderen Worten: dem Bildungsniveau kommt als selektivem Merkmal in Hinblick auf die Frage, ob eine Person verheiratet oder ledig ist, weniger Bedeutung zu als der Altersgruppenzugehörigkeit, dem Beobachtungszeitpunkt oder dem Geschlecht.

Übersicht 2: Modellanpassung und Devianzwerte für den individuellen Haushaltsmodus (Ledigenhaushalt versus Ehepaarhaushalt). Westdeutschland 1970, 1997; N = 237.115

Modell	Modellanpassung			Devianzaufklärung				
	G ²	DF	P	Pseudo-R ² (%) ^{a)}	G ²	DF	P	Pseudo-R ² (%)
A0. Unabhängigkeitsmodell^{b)}	40114,1	111	0,00	-				
A: Haupteffekte (RM1)								
A1= Basismodell: 1+B+G+A+J	5028,4	100	0,00	14,4				
A2 = A1 - B	7420,5	106	0,00	13,4	2392,1	6	0,00	1,0
A3 = A1 - G	8459,9	101	0,00	13,0	3431,5	1	0,00	1,4
A4 = A1 - A	21260,3	103	0,00	7,7	16231,9	3	0,00	6,7
A5 = A1 - J	17339,9	101	0,00	9,3	12311,5	1	0,00	5,1
B: Interaktionseffekte (RM2)								
B0 = A1+B*G+B*J+B*A +G*A+G*J+A*J	888,8	63	0,00	16,1				
B1 = B0 - B*G	2830,2	69	0,00	15,3	1941,5	6	0,00	0,8
B2 = B0 - B*A	967,4	81	0,00	16,0	78,7	18	0,00	0,1
B3 = B0 - G*A	1919,9	66	0,00	15,6	1031,2	3	0,00	0,5
B4 = B0 - B*J	973,3	69	0,00	16,0	84,5	6	0,00	0,1
B5 = B0 - G*J	1403,5	64	0,00	15,8	514,7	1	0,00	0,3
B6 = B0 - A*J	1763,0	66	0,00	15,7	874,3	3	0,00	0,4

a) McFaddens Pseudo-R²

b) Das empirisch maximal aufklärbare Pseudo-R² beträgt 16,4 Prozent. Allgemein deutet ein Pseudo-R² kleiner fünf Prozent auf einen eher geringen Zusammenhang. Werte über 20 Prozent auf starke Zusammenhänge hin (Andreß et al. 1997: 288f.; Hensher/Johnson 1981: 51).

Variablendefinition: B: Individuelles Bildungsniveau; G: Geschlecht; J: Beobachtungszeitpunkt; A: Altersgruppenzugehörigkeit

In den bisherigen Modellen wurde unterstellt, dass die Heiratsneigung zwar durch das Bildungsniveau, Alter, Geschlecht und den Beobachtungszeitpunkt beeinflusst wird, zwischen diesen Faktoren jedoch keine Interaktion besteht. Ausgehend von den oben ausgeführten theoretischen Überlegungen werden im Weiteren ausgewählte Interaktionseffekte in die Modellschätzung eingeführt. Mittels des Interaktionseffekts '*Bildungsniveau*Geschlecht*' wird die These der Existenz eines für Frauen negativen, für Männer positiven Bildungseffekts einer Überprüfung unterzogen (Brüderl/Diekmann 1997; Huinink 1995). Mit dem Interaktionseffekt '*Bildungsniveau*Alter*' wird berücksichtigt, dass hochqualifizierte Personen aufgrund der längeren Verweildauer in Bildungsinstitutionen typischerweise später

heiraten als weniger qualifizierte Personen. Über den Interaktionseffekt '*Geschlecht*Alter*' wird das geschlechtsspezifisch variierende Heiratsalter kontrolliert. Die These eines im Zeitverlauf zunehmenden Bildungseffekts wird über den Interaktionseffekt '*Bildungsniveau*Beobachtungszeitpunkt*' einer Überprüfung unterzogen. Wie oben ausgeführt, sollten die Modernisierungsprozesse nicht nur bei Frauen, sondern auch bei Männern den Anreiz zu einer Eheschließung im Zeitverlauf generell verringern, dieser Sachverhalt wird über den Interaktionseffekt '*Geschlecht*Beobachtungszeitpunkt*' abgebildet. Um die erwähnte Verschiebung des Heiratsalters im Zeitverlauf kontrollieren zu können, wird schliesslich noch der Interaktionseffekt '*Alter*Beobachtungszeitpunkt*' eingeführt.

Die Ergebnisse der Modellüberprüfungen sind in Übersicht 2 (Teiltabelle B) wiedergegeben. Wie aus dem Vergleich der Modelle B1 bis B6 hervorgeht, kommt der Interaktion zwischen *Bildungsniveau*Geschlecht* (Modell B1) die höchste Erklärungsleistung für die Zugehörigkeit zu einem der beiden Haushaltsmodi (Ehepaar- versus Ledigenhaushalt) zu. Eine gleichfalls hohe Erklärungsleistung geht von der Interaktion zwischen *Geschlecht*Alter* (Modell B3) aus, was wenig erstaunlich ist, da das Heiratsalter bekanntermaßen nach Geschlecht variiert.¹⁷ Die signifikante Interaktion zwischen *Alter*Beobachtungszeitpunkt* (Modell B6) deutet daraufhin, dass es in den letzten 30 Jahren zu einer Verschiebung des Heiratsalters gekommen ist. Auch dieser Befund ist - wie bereits erwähnt - bekannt. Ebenfalls von Bedeutung ist die Interaktion *Geschlecht*Beobachtungszeitpunkt* (Modell B5). Es ist daher davon auszugehen, dass sich die geschlechtsspezifischen Unterschiede im Zeitverlauf verändert haben. Im Vergleich hierzu kommt den Interaktionen *Bildung*Alter* (Modell B2) sowie *Bildung*Beobachtungszeitpunkt* (Modell B4) eine deutlich geringere Erklärungskraft zu. Mit anderen Worten: Es sind zwar alle Interaktionseffekte signifikant, die stärksten Einflüsse auf die Veränderungen im Zeitverlauf finden sich hierbei jedoch nicht in Bezug auf das Bildungsniveau, sondern im Hinblick auf Alter und Geschlecht. Demnach unterlag der Bildungseffekt im Zeitverlauf in Hinblick auf das Heiratsverhalten weniger Veränderungen als der Alters- und Geschlechtseffekt.

Für die folgende genauere Betrachtung des Heiratsverhaltens in Abhängigkeit von den oben angeführten Faktoren wird *Modell B0* herangezogen. Die Einzelkoeffizienten von Modell B0 sind in Tabelle A2 im Anhang dokumentiert. Die Interpretation der sich aus den Einzeleffekten ergebenden Muster und ihrer Veränderungen über die Zeit erfolgt anhand einer grafischen Darstellung zunächst für Frauen und anschliessend für Männer.

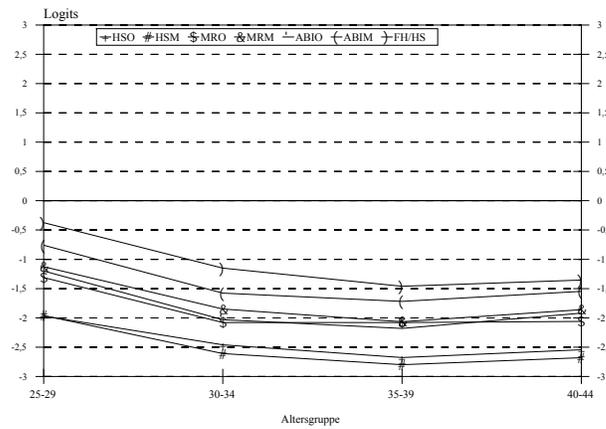
¹⁷ Nach Angaben des Statistischen Bundesamts (2000: 252) betrug das durchschnittliche Heiratsalter (Westdeutschland) im Jahr 1997 bei Männern 30,4 Jahre (1970: 25,6), bei Frauen 27,9 Jahre (1970: 23).

4.1.1 Frauen: Bildungsselektive Heiratsneigung 1970 und 1997 (Westdeutschland)

Grafik 1 zeigt die für Frauen zum Beobachtungszeitpunkt 1970 auf Basis von *Modell B0* (Übersicht 2) ermittelte logarithmierte Chance (Logits), verheiratet oder ledig zu sein, in Abhängigkeit vom Bildungsniveau und der Alterskohorte. Die Logits werden auf der y-Achse abgebildet. *Positive Werte* zeigen an, dass die *Chance ledig zu sein* überwiegt, bei *negativen Werten* ist die *Chance verheiratet zu sein* größer. Auf der x-Achse sind die Altersgruppen abgetragen. Die Linien repräsentieren die Zugehörigkeit zu den verschiedenen Bildungsgruppen.

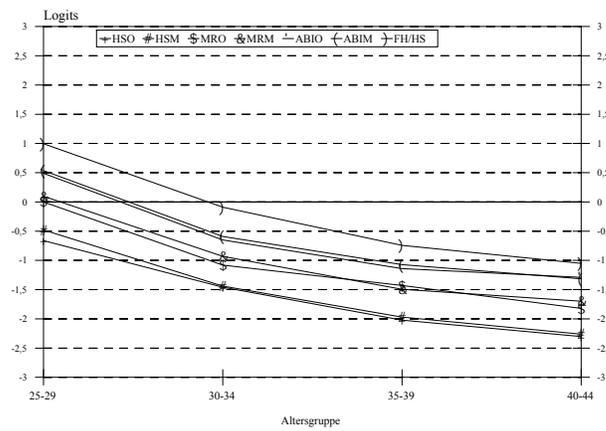
Wie aus Grafik 1 aufgrund der durchgängig negativen Logits erstens hervorgeht, war zum Beobachtungszeitpunkt 1970 die Chance, verheiratet zu sein, für Frauen insgesamt größer als die Chance, Single zu sein. Dieses Muster ist für alle betrachteten Bildungs- und Altersgruppen beobachtbar, wobei - wenig überraschend - die jüngeren Altersgruppen mit höherer Wahrscheinlichkeit ledig sind als die älteren Gruppen. Zweitens sprechen die Daten deutlich für einen Zusammenhang zwischen dem weiblichen Bildungsniveau und der Heiratsneigung: Je höher das Bildungsniveau, desto größer ist die Chance, ledig zu sein: Die höchste Ledigenchance findet sich über alle Altersgruppen hinweg bei den Akademikerinnen (FH/HS), gefolgt von Abiturientinnen mit Berufsausbildung (ABIM). Die geringste Ledigenchance weisen Hauptschulabsolventinnen mit oder ohne berufliche Qualifikation (HSO; HSM) auf. So haben die 40- bis 44-jährigen Frauen bei einem akademischen Abschluss (FH/HS) eine circa 4fach ($1/e^{-1,35}$), bei Abitur mit beruflicher Qualifikation (ABIM) eine fast 5fach ($1/e^{-1,54}$) höhere Chance verheiratet zu sein. In der Gruppe der 40- bis 44-jährigen beruflich qualifizierten Frauen mit Hauptschulabschluss (HSM) überwiegt die Chance, verheiratet zu sein hingegen um den Faktor 14 ($1/e^{-2,67}$). Zusammenfassend bestätigen die Ergebnisse damit zunächst den von der familienökonomischen Theorie postulierten negativen Bildungseffekt für Frauen. Dabei tritt dieser Effekt allerdings nicht erst, wie Beckers Ausführungen nahe legen, infolge der verstärkten Bildungspartizipation von Frauen auf, sondern bestand offensichtlich auch schon Anfang der 1970er Jahre.

Grafik 1: Effekte (Logits) des Bildungsniveaus von Frauen auf den individuellen Haushaltsmodus (Ledigen- vs. Ehepaarhaushalt) nach Altersgruppen (Westdeutschland, 1970)



Basis: VZ1970 (ZUMA-File); Modell B0 (Übersicht 2)

Grafik 2: Effekte (Logits) des Bildungsniveaus von Frauen auf den individuellen Haushaltsmodus (Ledigen- vs. Ehepaarhaushalt) nach Altersgruppen (Westdeutschland, 1997)



Basis: MZ1997 (ZUMA-File); Modell B0 (Übersicht 2)

Dieser Befund spricht für die oben formulierte These, nach welcher auch schon vor der Bildungsexpansion ein enger Zusammenhang zwischen dem Ausbildungsniveau von Frauen und ihrer Erwerbsbeteiligung zu finden ist, weshalb das Problem der mit zunehmender Qualifikation steigenden Opportunitätskosten auch hier schon zu erwarten wäre und sich in einer entsprechenden bildungsselektiven Heiratsneigung zeigen sollte.

In Grafik 2 sind die Befunde von Modell B0 für Frauen zum Beobachtungszeitpunkt 1997 abgebildet. Sofern die These eines im *Zeitverlauf zunehmenden Bildungseffekts* zutrifft, müsste sich der Abstand zwischen den oberen und mittleren bzw. unteren Bildungsgruppen 1997 im Vergleich zu 1970 vergrößert haben.

Bei Betrachtung der entsprechenden Effekte in Grafik 2 zeigt sich, dass diese These eines zunehmenden Bildungseffekts nicht zutrifft. Nach wie vor ist zwar ein deutlicher Bildungseffekt gegeben, aber die relative Differenz der Heiratswahrscheinlichkeiten der Akademikerinnen zu den restlichen Bildungsgruppen hat sich nicht in nennenswerter Weise vergrößert. Eine Veränderung des Bildungseffekts ist allerdings insofern zu beobachten, als sich die Heiratsneigung der Abiturientinnen ohne berufsqualifizierenden Abschluss (ABIO) an die der Abiturientinnen mit Berufsabschluss (ABIM) angepasst hat. Ende der 1990er Jahre verläuft der Bildungseffekt damit deutlicher als Anfang der 1970er Jahre entlang des dreigliedrigen allgemeinbildenden westdeutschen Schulsystems: die höchste Heiratswahrscheinlichkeit findet sich bei Hauptschulabsolventinnen (HSO; HSM), gefolgt von Frauen mit einem mittleren Bildungsabschluss (MRO; MRM) und Abiturientinnen (ABIO; ABIM).¹⁸ Die geringste Heiratswahrscheinlichkeit weisen Hochschulabsolventinnen (FH/HS) auf. Dies ist insofern ein etwas überraschender Befund, als die Erwerbschancen in Deutschland in einem hohen Maße mit der beruflichen Qualifikation verbunden sind. Vor diesem Hintergrund wäre nach der familienökonomischen Theorie zu erwarten gewesen, dass sich der Bildungseffekt eher an dem Kriterium der beruflichen Qualifikation als an der allgemeinbildenden Qualifikation orientiert. Welche Faktoren zur Herausbildung dieses Musters führen, muss einer weiterführenden Analyse vorbehalten sein. Stattdessen soll das Augenmerk auf eine andere, wesentlich wichtigere Entwicklung gerichtet werden: Die im Vergleich zu 1970 starke Niveaushiftung in Richtung Ledige über alle Bildungs- und Altersgruppen hinweg. Zwar ist diese Entwicklung bei den jüngeren Altersgruppen (möglicherweise als Resultat des ansteigenden Heiratsal-

18 In den oben erwähnten früheren Studien zur bildungsselektiven Heiratsneigung in Deutschland ist ein solcher Effekt nicht erkennbar, da das Ausbildungsniveau in aller Regel nicht kategorial, sondern metrisch (d.h. über die Bildungsjahre) operationalisiert wurde.

ters) am stärksten ausgeprägt,¹⁹ sie findet sich aber auch bei den beiden älteren Kohorten. Mit anderen Worten, der in soziodemographischen Studien beobachtete Anstieg der Ledigenquoten ist unseren Befunden zufolge nicht das Resultat einer im Zeitverlauf *zunehmend* bildungsselektiven Heiratsneigung bzw. Ausdruck einer Verhaltensänderung von hochqualifizierten Frauen, sondern beruht auf einer für alle Bildungsgruppen in ähnlicher Weise zu beobachtenden rückläufigen Heiratsneigung.

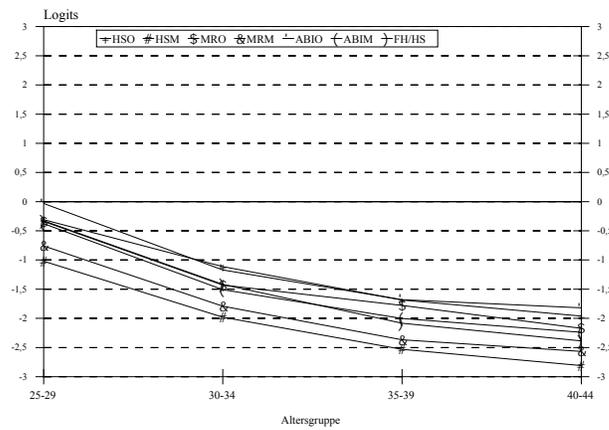
Als vorläufiges Fazit ist festzuhalten, dass die Befunde einerseits sehr klar eine bildungsselektive Heiratsneigung von Frauen aufzeigen, welche sich in Form einer mit zunehmenden Bildungsniveau sinkenden Heiratswahrscheinlichkeit zeigt. Andererseits hat sich dieser Bildungseffekt innerhalb der beobachteten Zeitspanne nicht verstärkt. Der Vergleich zwischen 1970 und 1997 deutet vielmehr auf eine Art Fahrstuhleffekt hin, bei welchem für alle Bildungsgruppen gleichermaßen die Wahrscheinlichkeit, ledig zu sein, angestiegen ist.

4.1.2 Männer: Bildungsselektive Heiratsneigung 1970 und 1997 (Westdeutschland)

Grafik 3 bildet analog zu oben die Befunde für Männer (Modell B0, Übersicht 2) zum Erhebungszeitpunkt 1970 ab. Es wird deutlich, dass die Heiratsneigung von Männern im Vergleich zu Frauen eine erheblich geringer ausgeprägte bildungsspezifische Variation und zugleich ein anderes Muster aufweist. Dieses Muster folgt allerdings nicht der oben für Männer formulierten These eines positiven Bildungseffekts, d.h. einer mit der Qualifikation zunehmenden Heiratswahrscheinlichkeit. Stattdessen findet sich die höchste Heiratswahrscheinlichkeit bei Hauptschulabsolventen mit Berufsqualifikation (HSM) sowie Männern mit mittlerem berufsqualifizierenden Abschluss (MRM). Akademiker (FH/HS) und Abiturienten mit Berufsabschluss (ABIM) nehmen hingegen nur eine mittlere Position ein. Die geringste Heiratswahrscheinlichkeit weisen Hauptschulabsolventen und Abiturienten jeweils ohne berufliche Qualifikation (HSO; ABIO) auf. Der letztere Befund steht in Übereinstimmung mit den theoretischen Annahmen und kann darüber erklärt werden, dass Männer, die einen gewissen qualifikatorischen Minimalstandard nicht erreichen und ein dementsprechend geringes Einkommenspotenzial aufweisen, unter ökonomischen Kriterien eine eher ungünstige Position auf dem Heiratsmarkt innehaben.

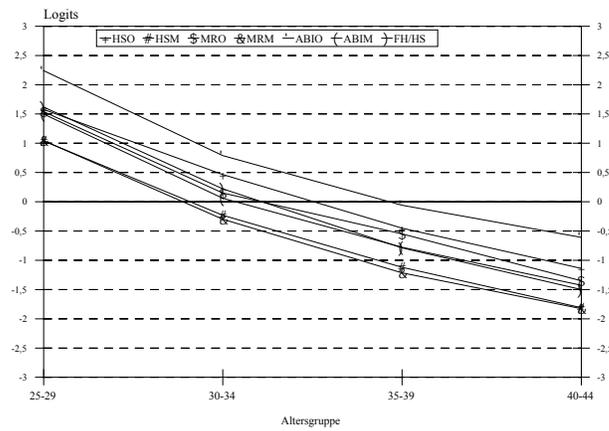
¹⁹ Inwieweit es sich hierbei tatsächlich um einen Alterseffekt handelt, d.h. in welchem Umfang die jüngeren Kohorten in einem höheren Alter noch heiraten werden, wird sich in ein paar Jahren zeigen.

Grafik 3: Effekte (Logits) des Bildungsniveaus von Männern auf den individuellen Haushaltsmodus (Ledigen- vs. Ehepaarhaushalt) nach Altersgruppen (Westdeutschland, 1970)



Basis: VZ1970 (ZUMA-File); Modell B0 (Übersicht 2)

Grafik 4: Effekte (Logits) des Bildungsniveaus von Männern auf den individuellen Haushaltsmodus (Ledigen- vs. Ehepaarhaushalt) nach Altersgruppen (Westdeutschland, 1997)



Basis: MZ1997 (ZUMA-File); Modell B0 (Übersicht 2)

Betrachtet man im Weiteren die Befunde für 1997 (Grafik 4), so zeigt sich, dass die Heiratswahrscheinlichkeit bei Männern wesentlich deutlicher als bei Frauen über alle Alters- und Bildungsgruppen rückläufig ist. Zugleich hat sich die bildungsselektive Heiratsneigung etwas vergrößert, ohne allerdings das für Frauen beobachtete Niveau zu erreichen. Bemerkenswert ist dabei die Gruppe der Abiturienten ohne Berufsausbildung (ABIO), bei welcher die Heiratsneigung erheblich stärker abgenommen hat als bei den anderen Bildungsgruppen. Welche Mechanismen hinter dieser Entwicklung stehen, ist auf Basis der hier vorhandenen Daten nicht zu klären. Die auf den ersten Blick naheliegende Vermutung, dass sich diese Gruppe zum Teil noch in der Ausbildungsphase befindet, ist zum einen insofern nicht plausibel, als Personen in Ausbildung nicht in die Analyse aufgenommen wurden. Gegen diese Annahme spricht auch, dass die Abweichungen bei allen Altersgruppen zu beobachten ist. Auch die für 1970 formulierte Vermutung einer ungünstigen Position dieser Gruppe auf dem Heiratsmarkt ist wenig schlüssig, da dann ähnliche Effekte für die anderen Bildungsgruppen ohne berufliche Qualifikation (MRO; HSO) zu erwarten wären.

Insgesamt ist festzuhalten, dass die empirisch beobachtete Bildungsselektivität im Heiratsverhalten von Männern von den auf Basis der Familienökonomie oben formulierten Thesen in Teilen abweicht und diesbezüglich weiterer Erklärungsbedarf sowohl in theoretischer wie auch in empirischer Hinsicht besteht. Hiervon unabhängig lassen die Befunde jedoch auch erkennen, dass die wichtigste Veränderung zwischen Anfang der 1970er und Ende der 1990er Jahre auch bei den Männern *nicht* in bezug auf den *Bildungseffekt* zu beobachten ist, sondern in der *allgemeinen Niveaushiftung zugunsten der Ledigen* über alle Bildungs- und Altersgruppen hinweg besteht.

4.1.3 Bildungsselektive Heiratsneigung: Ostdeutschland - Westdeutschland im Vergleich

Abschliessend soll die These der bildungsselektiven Heiratsneigung im Ost-Westvergleich betrachtet werden. Wie oben ausgeführt, konzentriert sich das familienökonomische Erklärungsmodell des Heiratsverhaltens auf individuelle Entscheidungsprozesse. Der mögliche Einfluss von im Zuge gesellschaftlicher Modernisierungsprozesse sich ändernden sozialpolitischen Rahmenbedingungen oder auch normativen Erwartungen auf das Heiratsverhalten und die Familiengründungsprozesse bleibt im Erklärungsmodell ausgeblendet. Dies ist aber nicht gleichbedeutend damit, dass diesen Faktoren keine Bedeutung für individuelles Handeln zukommt. Wäre dem so, könnte man in Zukunft u.a. auf familienpolitische Massnahmen verzichten, da diese ohnehin folgenlos bleiben würden. Für eine erste explorative Betrachtung des Einflusses von gesellschaftspolitischen und sozialen Randbedingungen

auf das Heiratsverhalten bietet sich ein Vergleich zwischen Ost- und Westdeutschland an. Denn ähnlich wie in Westdeutschland verläuft die Arbeitsteilung im Haushalt auch in Ostdeutschland im Wesentlichen entlang der traditionellen geschlechtsspezifischen Rollenzuweisung (Hartmann 1998). Im Unterschied zu Westdeutschland wurde aber die Erwerbstätigkeit von Frauen und vor allem von Müttern in der damaligen DDR (nicht zuletzt aus volkswirtschaftlichen Gründen) wesentlich früher und stärker gesellschaftlich akzeptiert als im Westen.²⁰ Dem aus der Doppelbelastung durch Haus- und Erwerbsarbeit resultierenden Zielkonflikt, welcher der familienökonomischen Theorie zufolge als Ursache für die bildungsselektive Heiratsneigung von Frauen anzusehen ist, wurde durch eine Vielzahl von sozialpolitischen Massnahmen entgegen gesteuert. Hierzu gehörten eine Reihe von Privilegien (z.B. verkürzte Arbeitszeit, Haushaltstag, Babyjahr, Erziehungsurlaub) für erwerbstätige Mütter als auch die Bereitstellung einer ausreichenden Zahl von Kinderbetreuungsstätten. Darüber hinaus wurde die Eheschließung durch selektive Anreize (z.B. Anrecht auf eine eigene Wohnung bei Eheschließung; Gewährung günstiger Kredite für junge Ehepaare) unterstützt. Es ist anzunehmen, dass diese umfangreichen sozialpolitischen Massnahmen die Opportunitätskosten bei Eheschließung und Familiengründung für hochqualifizierte Frauen in der früheren DDR im Vergleich zu Westdeutschland reduziert haben (vgl. Brüderl/Diekman 1994: 60; 1997: 9). Empirisch sollte sich dies darin zeigen, dass die älteren ostdeutschen Frauenkohorten, deren Ausbildungs- und Familiengründungsphase im Wesentlichen vor der Wiedervereinigung abgeschlossen war, eine im Vergleich zu den Westkohorten deutlich geringere bildungsselektive Heiratsneigung aufweisen.

Zur Überprüfung dieser These greifen wir auf das obige Modell (Übersicht 2) zurück, wobei die Variable Beobachtungszeitpunkt durch die Variable Ost-West ersetzt wird.²¹ Übersicht 3 dokumentiert die Ergebnisse der binomialen logistischen Regressionsmodelle.

Wie aus dem Vergleich der Modelle A2 bis A5 ersichtlich, leistet die Ost-West Variable (Modell A5) zwar einen signifikanten Erklärungsbeitrag zum Haushaltsmodus, allerdings ist dieser Beitrag im Vergleich zu den anderen Merkmalen gering. Die größte Erklärungskraft geht von der Altersgruppenzugehörigkeit aus (Modell A4), gefolgt von Geschlecht (Modell A3) und Bildungsniveau (Modell A2).

²⁰ Dieses Muster zeigt sich auch in der Gegenwart in Form der in Ostdeutschland im Vergleich zu Westdeutschland modernen Einstellungen zur Rolle der Frau (Kurz 1998: 13).

²¹ Einen Überblick der geschlechtsspezifischen Bildungs- und Altersverteilungen nach Haushaltsmodus für Ostdeutschland gibt Tabelle A1 im Anhang.

Übersicht 3: Modellanpassung und Devianzwerte für den individuellen Haushaltsmodus (Ledigenhaushalt versus Ehepaarhaushalt). Vergleich Ostdeutschland - Westdeutschland 1997; N=112.003

Modell	Modellanpassung				Devianzaufklärung			
	G ²	DF	P	Pseudo-R ² (%) ^{a)}	G ²	DF	P	Pseudo-R ² (%)
A0. Unabhängigkeitsmodell^{b)}	21480,9	111	0,00	-				
A: Haupteffekte (RM1)								
A1= Basismodell: 1+B+G+A+O	1068,8	100	0,00	14,3				
A2 = A1 - B	2222,3	106	0,00	13,5	1153,5	6	0,00	0,8
A3 = A1 - G	3901,6	101	0,00	12,3	2832,8	1	0,00	2,0
A4 = A1 - A	17998,7	103	0,00	2,4	16929,9	3	0,00	11,9
A5 = A1 - O	1111,9	101	0,00	14,3	43,1	1	0,00	0,0
B: Interaktionseffekte (RM2)								
B0 = A1+B*G+B*O+B*A +G*A+G*O+A*O	86,5	63	0,00	15,0				
B1 = B0 - B*G	460,6	69	0,00	14,7	374,1	6	0,00	0,3
B2 = B0 - B*A	174,1	81	0,00	14,9	87,6	18	0,00	0,1
B3 = B0 - G*A	146,2	66	0,00	15,0	59,7	3	0,00	0,0
B4 = B0 - B*O	372,0	69	0,00	14,8	285,5	6	0,00	0,2
B5 = B0 - G*O	100,3	64	0,00	15,0	13,8	1	0,00	0,0
B6 = B0 - A*O	162,7	66	0,00	14,9	76,2	3	0,00	0,1

a) McFaddens Pseudo-R²

b) Das empirisch maximal aufklärbare Pseudo-R² beträgt 15,1 Prozent.

Variablendefinition:

B: Individuelles Bildungsniveau; G: Geschlecht; O: Ost-West-Variable;

A: Altersgruppenzugehörigkeit

In den Modellen B1 bis B6 werden ausgewählte Interaktionseffekte getestet. Berücksichtigt werden die bereits unter 4.1 diskutierten Interaktionseffekte '*Bildung*Geschlecht*' (Modell B1), '*Bildung*Alter*' (Modell B2) sowie '*Geschlecht*Alter*' (Modell B3). Der oben postulierten These einer in Ostdeutschland geringeren bildungsselektiven Heiratsneigung wird über die Einführung des Interaktionseffekts '*Bildung*Ost-West*' (Modell B4) Rechnung getragen. Für die Überprüfung geschlechtsspezifischer Variationen im Heiratsverhalten zwischen Ost- und Westdeutschland wird der Effekt '*Geschlecht*Ost-West*' (Modell B5) einbezogen. Mit dem Interaktionseffekt '*Alter*Ost-*

West' (Modell B6) wird berücksichtigt, dass in Ostdeutschland in einem jüngeren Alter geheiratet wurde als in Westdeutschland.²² Insgesamt führen die verschiedenen Interaktionseffekte zu einem leichten Anstieg der Erklärungskraft auf 15 Prozent (Modell B0), wobei diese Erhöhung im Wesentlichen durch die Interaktionen '*Bildung*Geschlecht*' sowie '*Bildung*Ost-West*' bewirkt wird. Wenngleich diese Befunde damit eine sparsamere Modellierung nahelegen, greifen wir für die folgende Darstellung der bildungsselektiven Heiratsneigung in Ostdeutschland auf Modell B0 zurück, da sich hierdurch der Vergleich zu den obigen Ausführungen für Westdeutschland systematischer gestalten lässt.²³

Grafik 5 bildet für ostdeutsche Frauen (Beobachtungszeitpunkt 1997) die auf Basis von Modell B0²⁴ ermittelten logarithmierten Chancen (Logits), verheiratet oder ledig zu sein, in Abhängigkeit vom Bildungsniveau und der Altersgruppe ab. In überraschender Deutlichkeit weicht das für Ostdeutschland beobachtete Muster von den entsprechenden Ergebnissen für Westdeutschland (vgl. Grafik 2) ab. So zeigt sich, dass der Effekt des Bildungsniveaus auf das Heiratsverhalten ostdeutscher Frauen erheblich schwächer ausgeprägt ist als im Westen: die Heiratswahrscheinlichkeit zwischen den einzelnen Bildungsgruppen unterscheidet sich in Ostdeutschland nur marginal. Eine Ausnahme hiervon stellt lediglich die Gruppe der beruflich nicht qualifizierten Hochschulabsolventinnen (HSO) dar, wobei es sich hier allerdings um eine insgesamt nur sehr kleine Gruppe handelt.²⁵

Darüber hinaus ist im Vergleich zu Westdeutschland eine Umkehrung der Wirkungsrichtung des Bildungsniveaus zu beobachten. Die am geringsten qualifizierten Frauen (HSO) weisen die niedrigste, Frauen mit mittlerem berufsqualifizierenden Abschluss (MRM) die höchste Heiratswahrscheinlichkeit auf. Wobei diese Unterschiede allerdings insgesamt so gering sind, dass sie nicht überinterpretiert werden sollten. Oben wurde argumentiert,

22 Seit der Wiedervereinigung hat sich das durchschnittliche Heiratsalter in Ostdeutschland an westdeutsche Verhältnisse angenähert. Im Jahr 1989 lag das durchschnittliche Heiratsalter in Ostdeutschland für Männer bei 25,3 Jahren (West: 28,2) für Frauen bei 23,2 Jahren (West: 25,7) (Statistisches Bundesamt 1992: 75). 1997 hat sich das durchschnittliche Heiratsalter in Ostdeutschland für Männer auf 29,3 Jahre (West: 30,4) für Frauen bei 27 Jahren (West: 27,9) erhöht (Statistisches Bundesamt 2000: 252).

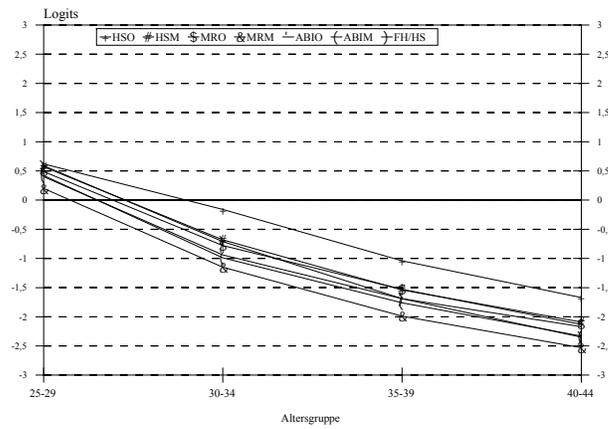
23 Die auf Basis des Zeitvergleichs für Westdeutschland ermittelten Koeffizienten für 1997 weichen von den entsprechenden Koeffizienten des Ost-West-Vergleichs etwas ab. Dies ist durch Einbeziehung unterschiedlicher Interaktionseffekte bedingt. Auf das für Westdeutschland dargestellte Basismuster und die Interpretation hat dies jedoch keinen Einfluss, weshalb für den Ost-West-Vergleich die Grafik 2 (Frauen; Westdeutschland 1997) bzw. Grafik 4 (Männer; Westdeutschland 1997) herangezogen wird.

24 Die Koeffizienten für Modell B0 aus Übersicht 3 sind in Tabelle A3 im Anhang dokumentiert.

25 Vgl. Tabelle A1 im Anhang.

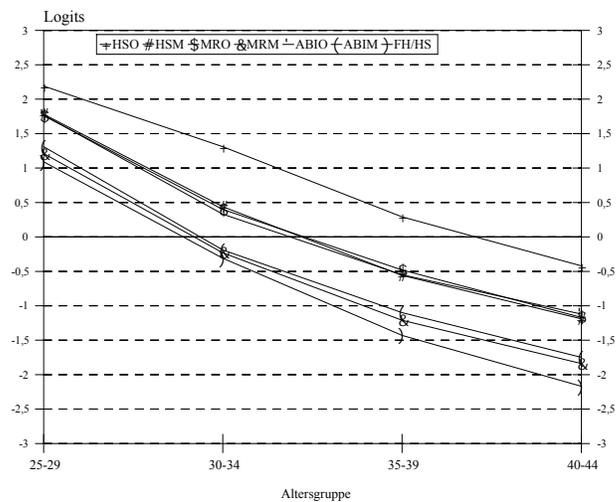
dass institutionelle und soziale Randbedingungen in Bezug auf das Heiratsverhalten als wesentliche Einflussfaktoren anzusehen sind.

Grafik 5: Effekte (Logits) des Bildungsniveaus von Frauen auf den individuellen Haushaltsmodus (Ledigen- vs. Ehepaarhaushalt) nach Altersgruppen (Ostdeutschland, 1997)



Basis: MZ1997 (ZUMA-File); Modell B0 (Übersicht 3)

Grafik 6: Effekte (Logits) des Bildungsniveaus von Männern auf den individuellen Haushaltsmodus (Ledigen- vs. Ehepaarhaushalt) nach Altersgruppen (Ostdeutschland, 1997)



Basis: MZ1997 (ZUMA-File); Modell B0 (Übersicht 3)

Die Befunde für die ostdeutschen Frauen unterstützen diese These. Dennoch sind die für Ost- und Westdeutschland auch noch Jahre nach der Vereinigung zu beobachtenden Unterschiede doch überraschend. Insbesondere der Vergleich der jüngsten Altersgruppe (25- bis 29-Jährige), bei welcher sich für Westdeutschland das typische bildungsselektive Muster zeigt, in Ostdeutschland hingegen die bildungsspezifischen Unterschiede noch geringer sind als bei den älteren Kohorten, macht deutlich, dass monokausale Erklärungen zu kurz greifen. Für eine präzisere Einschätzung des von gesellschaftlichen Randbedingungen ausgehenden Einflusses auf das Heiratsverhalten wäre dabei eine international vergleichende Untersuchung weiterführend.

Abgeschlossen sind in Grafik 6 die Effekte aus Modell B0 für ostdeutsche Männer abgebildet. Im Unterschied zu den Frauen zeigen sich für Männer in Ostdeutschland deutliche bildungsspezifische Effekte in Form einer mit dem Qualifikationsniveau ansteigenden Heiratsneigung: Die geringste Heiratswahrscheinlichkeit weisen die am wenigsten qualifizierten Männer (HSO) auf, gefolgt von einer Gruppe, die sich aus Hauptschulabsolventen mit Berufsausbildung (HSM) und beruflich nicht qualifizierten Männern mit mittlerem Bildungsabschluss oder Abitur (MRO, ABIO) zusammensetzt. Die höchste Wahrscheinlichkeit, verheiratet zu sein, findet sich bei Männern mit mittlerem Abschluss oder Abitur und jeweils beruflicher Qualifikation sowie bei den Akademikern (MRM, ABIM, FH/HS). Mit anderen Worten: die auf Basis der familienökonomischen Theorie oben formulierte These eines positiven Bildungseffekts für Männer, die in Westdeutschland allenfalls in abgeschwächter Form (vgl. Grafik 4) zutrifft, wird durch die Ergebnisse für ostdeutsche Männer unterstützt.

Über die hinter diesen Befunden stehenden Wirkungsmechanismen könnte an dieser Stelle nur spekuliert werden. Unseres Erachtens lassen die berichteten Ergebnisse zum jetzigen Zeitpunkt nur den Schluss zu, dass offensichtliche Unterschiede im bildungs- und geschlechtsspezifischen Heiratsverhalten zwischen Ost- und Westdeutschland bestehen, die nicht in einfacher Weise mittels der familienökonomischen Theorie erklärt werden können. Eine weiterführende Analyse muß von alternativen Erklärungsmodellen ausgehen, bei welchen Bildung nicht nur als Humankapital, sondern auch als soziales und kulturelles Kapital berücksichtigt wird und in welchen darüber hinaus auch die jeweils vorzufindenden gesellschaftspolitischen Randbedingungen theoretisch fundiert integriert werden können.

5. Zusammenfassung

Im Zentrum des vorliegenden Beitrags stand die Frage, ob und in welchem Umfang das Heiratsverhalten durch das individuelle Bildungsniveau beeinflusst ist. Die auf Basis der familienökonomischen Theorie abgeleiteten Hypothesen wurden durch die empirischen Analysen dabei nur zum Teil unterstützt. Die Ergebnisse sprechen in eindeutiger Weise für ein ausgeprägt bildungsselektives Heiratsverhalten von Frauen in Westdeutschland: je höher die Qualifikation, desto geringer die Heiratswahrscheinlichkeit. Allerdings, und dies erscheint uns als der wichtigste Befund, hat sich die Bildungsselektivität in den letzten 30 Jahren nicht vergrößert. Der relative Abstand zwischen den Heiratswahrscheinlichkeiten von hochqualifizierten Frauen und jenen der anderen Bildungsgruppen ist nahezu konstant geblieben. Vielmehr ist die Entwicklung im Zeitverlauf durch einen für alle Bildungsgruppen gleichermaßen beobachtbaren Rückgang der Heiratsneigung gekennzeichnet. Insgesamt sprechen diese Ergebnisse deshalb dafür, dass der vielfach konstatierte Wandel im Heiratsverhalten und in den Familiengründungsprozessen nicht primär durch veränderte Verhaltensmuster hochqualifizierter Frauen bedingt ist, sondern sich eher zusammensetzt aus (1.) einer unabhängig vom Bildungsniveau sinkenden Heiratsneigung von Frauen und (2.) einem Kompositionseffekt: In den 70er Jahren war die Gruppe der hochqualifizierten Frauen vergleichsweise klein, deshalb wurde ihr 'abweichendes' Verhalten in bezug auf Heiratsneigung, Familiengründung und Erwerbsverhalten von der Öffentlichkeit und der Politik kaum wahrgenommen. Mit der Bildungsexpansion ist die Zahl der Hochqualifizierten stark angestiegen, die der Geringqualifizierten erheblich gesunken (vgl. Tabelle A1 im Anhang), damit kommt dem 'untypischen' Verhalten von hochqualifizierten Frauen in der Gegenwart ein sehr viel stärkeres Gewicht zu, als dies noch Anfang der 70er Jahre der Fall war.

Weniger eindeutig sind die Befunde in Hinblick auf das bildungsselektive Heiratsverhalten der westdeutschen Männer. Im Vergleich zu den Frauen zeigen sich nur schwach ausgeprägte Bildungseffekte, die sich im Zeitverlauf etwas verstärken. In der Tendenz deuten die Analysen hierbei darauf hin, dass sich die Heiratswahrscheinlichkeit von Männern beim Vorliegen eines berufsqualifizierenden Abschlusses erhöht. Ähnlich wie für Frauen basiert auch bei Männern die maßgebliche Veränderung nicht auf einer sich wandelnden Bildungsselektivität, sondern auf dem in allen Bildungsgruppen feststellbaren Rückgang der Heiratswahrscheinlichkeit.

Als weiterer Befund der Analysen ist festzuhalten, dass die Strukturen in Ost- und Westdeutschland im Hinblick auf die bildungsselektive Heiratsneigung differieren. In den neuen Bundesländern ist der Einfluss von Bildung auf das Heiratsverhalten von Frauen deutlich geringer als in den alten Bundesländern. Darüber hinaus deutet sich ein anderes

Muster an: Während in Westdeutschland eine hohe Qualifikation die Heiratswahrscheinlichkeit verringert, sind es in Ostdeutschland die am niedrigsten qualifizierten Frauen, die am häufigsten ledig bleiben. Für ostdeutsche Männer wiederum steigt die Wahrscheinlichkeit, verheiratet zu sein, mit dem Qualifikationsniveau.

Die Unterschiede zwischen Ost- und Westdeutschland sind unseres Erachtens ein wichtiger Indikator für den Einfluss von gesellschaftspolitischen Rahmenbedingungen auf das Heiratsverhalten. Allerdings bleibt zu fragen, weshalb die jeweiligen Randbedingungen in Westdeutschland bei Frauen stärkere Bildungseffekte hervorrufen als bei Männern, während sie in Ostdeutschland stärkere Effekte für Männer als für Frauen erzeugen. Dies ist gleichbedeutend mit der Frage, welche sozialen Wirkungsmechanismen im Einzelnen hinter der Herausbildung einer bildungsselektiven Heiratsneigung stehen. Auf Basis der hier herangezogenen Daten können diese Fragen sicherlich nicht beantwortet werden. Gleichwohl gehen wir davon aus, dass bei steigenden Ledigenquoten diese Fragen zukünftig verstärkt in Vordergrund rücken.

Korrespondenzadresse

Dr. Heike Wirth
ZUMA
Postfach 12 21 55
68 072 Mannheim
email: wirth@zuma-mannheim.de

Literatur

- Andreß, H.-J./Hagenaars, J.A./Kühnel, S., 1997: Analyse von Tabellen und kategorialen Daten. Berlin: Springer.
- Becker, G.S., 1981: A treatise on the family. Cambridge: Harvard University Press.
- Becker, G.S., 1991: A Treatise on the Family. Cambridge: Harvard University Press.
- Ben-Porath, Y., 1982: Economics and the Family - Match or Mismatch? A Review of Becker's A Treatise on the Family. *Journal of Economic Literature* 20: 52-64.
- Blossfeld, H.-P., 1995: Changes in the Process of Family Formation and Women's Growing Economic Independence: A Comparison of Nine Countries. S. 3-34 in: Blossfeld, H.-P. (Hrsg.), *The New Role of Women. Family Formation in Modern Societies*. Boulder: Westview Press.
- Blossfeld, H.-P./Huinink, J., 1989: Die Verbesserung der Bildungs- und Berufschancen von Frauen und ihr Einfluss auf den Prozess der Familienbildung. *Zeitschrift für Bevölkerungswissenschaft* 4: 383-404.

- Blossfeld, H.-P./Huinink, J./Rohwer, G., 1993: Wirkt sich das steigende Bildungsniveau der Frauen tatsächlich negativ auf den Prozess der Familienbildung aus? S. 216-233 in: Diekmann, A./Weick, S. (Hrsg.), *Der Familienzyklus als sozialer Prozess. Bevölkerungssoziologische Untersuchungen mit den Methoden der Ereignisanalyse*. Berlin: Duncker&Humblot.
- Blossfeld, H.-P./Jaenichen, U., 1990: Bildungsexpansion und Familienbildung. *Soziale Welt* 41: 454-476.
- Blossfeld, H.-P./Timm, A., 1997: Das Bildungssystem als Heiratsmarkt. Eine Längsschnittanalyse der Wahl von Heiratspartnern im Lebenslauf. SFB 186 Arbeitspapier Nr. 43: Bremen.
- Braun, M./Müller, W., 1997: Measurement of Education in Comparative Research. *Comparative Social Research* 16: 163-201.
- Brauns, H./Steinmann, S., 1999: Educational Reform in France, West-Germany, and the United Kingdom. Updating the CASMIN Educational Classification. *ZUMA-Nachrichten* 44: 7-44.
- Brüderl, J./Diekmann, A., 1994: Bildung, Geburtskohorte und Heiratsalter. Eine vergleichende Untersuchung des Heiratsverhaltens in Westdeutschland, Ostdeutschland und den Vereinigten Staaten. *Zeitschrift für Soziologie* 23, 1: 56-73.
- Brüderl, J./Diekmann, A., 1997: Education and Marriage. A comparative study. Unveröffentlichtes Arbeitspaier.
- Brüderl, J./Klein, T., 1993: Bildung und Familiengründungsprozesse deutscher Frauen: Humankapital- und Institutioneneffekt. S.194-215 in: Diekmann, A./Weick, S. (Hrsg.), *Der Familienzyklus als sozialer Prozess. Bevölkerungssoziologische Untersuchungen mit den Methoden der Ereignisanalyse*. Berlin: Duncker & Humblot.
- Burkart, G., 1997: *Lebensphasen Liebespaare*. Opladen: Leske+Budrich.
- Dorbritz, J./Gärtner, K., 1995: Die demographische Bedeutung des Familienstandes. Forschungsbericht. Stuttgart: Kohlhammer
- Dörr, G./Glatzer, W., 1995: Haushaltsproduktion und Wohlfahrtsproduktion - Strukturwandlungen und Zukunftsperspektiven. S. 515-532 in: Nauck, B./Onnen-Isemann, C. (Hrsg.), *Familie im Brennpunkt von Wissenschaft und Forschung*. Neuwied: Luchterhand.
- Handl, J., 1978: Ausmaß und Determinanten der Erwerbsbeteiligung von Frauen. S.189-256 in: Franke, H./Kaiser, M./Nuthmann, R./Stegmann, H. (Hrsg.), *Probleme bei der Konstruktion sozioökonomischer Modelle*. BeitrAB31: Nürnberg, IAB.
- Hannan, M.T., 1982: Families, Markets and Social Structures: An Essay on Becker's A Treatise on the Family. *Journal of Economic Literature* 20: 65-72.
- Hartmann, P., 1998: Arbeitsteilung im Haushalt. S.139-176 in: Mohler, P.Ph./Braun, M. (Hrsg.), *Soziale Ungleichheit in Deutschland. Blickpunkt Gesellschaft* 4. Opladen: Westdeutscher Verlag.
- Hensher, D.A./Johnson, L.W., 1981: *Applied Discrete-Choice Modelling*. London: Croom Helm.

- Höhn, C./Dorbritz, J., 1995: Zwischen Individualisierung und Institutionalisierung - Familiendemographische Trends im vereinten Deutschland. S.149-176 in: Nauck, B./Onnen-Isemann, C. (Hrsg.), Familie im Brennpunkt von Wissenschaft und Forschung. Neuwied: Luchterhand.
- Huinink, J., 1989: Ausbildung, Erwerbsbeteiligung von Frauen und Familienbildung im Kohortenvergleich. S.134-158 in: Wagner, G./Ott, N./Hoffmann-Nowotny, H.-J. (Hrsg.), Familienbildung und Erwerbstätigkeit im demographischen Wandel. Berlin: Springer.
- Huinink, J., 1995: Education, Work and Family Patterns of Men: The Case of West Germany. S.247-262 in: Blossfeld, H.-P. (Hrsg.), The New Role of Women. Family Formation in Modern Societies. Boulder: Westview Press.
- König, W./Lüttinger, P./Müller, W., 1987: Eine vergleichende Analyse der Entwicklung und Struktur von Bildungssystemen. Methodologische Grundlagen und Konstruktion einer vergleichbaren Bildungsskala. CASMIN-Projekt Arbeitspapier Nr. 12: Universität Mannheim.
- Kurz, K., 1998: Hausfrau oder Berufsfrau? Einstellungen zur Rolle der Frau in Ost- und Westdeutschland. S.173-220 in: Braun, M./Mohler, P.Ph. (Hrsg.), Soziale Ungleichheit in Deutschland. Opladen: Westdeutscher Verlag.
- Lechert, Y./Staudenmaier, B./Schmidt, S., 2000: Mikrozensus 1997: Dokumentation und Datenaufbereitung. ZUMA-Technischer Bericht Nr. 2000/13.
- Lüttinger, P./Riede, T., 1997: Der Mikrozensus. Amtliche Daten für die Sozialforschung. ZUMA-Nachrichten 41: 19-43.
- Nave-Herz, R.Hg., 1988: Wandel und Kontinuität der Familie in der Bundesrepublik Deutschland. Stuttgart: Enke.
- Oppenheim Mason, K./Jensen, A.-M., 1995: Gender and family change in industrialized countries. Oxford: Clarendon Press.
- Oppenheimer, V.K., 1988: A Theory of Marriage Timing. American Journal of Sociology 94, 1: 563-591.
- Oppenheimer, V.K., 1995: The role of Women's economic independence in Marriage formation: A sceptic response to Annemette Sorensen's Remarks. S.236-241 in: Blossfeld, H.-P. (Hrsg.), The New Role of Women. Family Formation in Modern Societies. Boulder: Westview Press.
- Oppenheimer, V.K., 1997: Women's employment and the gain to marriage. The specialization and trading model. Annual Review of Sociology 23: 431-453.
- Oppenheimer, V.K./Blossfeld, H.-P. /Wackerow, A., 1995: Country-Specific Studies on the Trend in Family Formation and the New Role of Women: United States of America. S.150-173 in: Blossfeld, H.-P. (Hrsg.), The New Role of Women. Family Formation in Modern Societies. Boulder: Westview Press.
- Oppenheimer, V.K./Lew, V., 1995: American Marriage Formation in the 1980s: How important was Women's Economic Independences? S.105-138 in: Oppenheim Mason,

- K./Jensen, A.-M. (Hrsg.): Gender and family change in industrialized countries. Oxford: Clarendon Press.
- Ostner, I., 1999: Ehe oder Familie - Konvention oder Sonderfall. Ursachen, Probleme und Perspektiven des Wandels der Lebensformen. Zeitschrift für Familienforschung 11: 1: 32-51.
- Ott, N., 1998: Der familienökonomische Ansatz von Gary S. Becker. S.63-90 in: Pies, I./Leschke, M. (Hrsg.), Gary Beckers ökonomischer Imperialismus. Tübingen: Mohr Siebeck.
- Pies, I./Leschke Martin, 1998: Gary Beckers ökonomischer Imperialismus. Tübingen: Mohr Siebeck.
- Röhler, H./Steinbach, A./Huinink, J., 2000: Hausarbeit in Partnerschaften. Zeitschrift für Familienforschung 2: 21-53.
- Schimpl-Neimanns, B., 1998: Analysemöglichkeiten des Mikrozensus. ZUMA-Nachrichten 42: 91-119.
- Schimpl-Neimanns, B., 2002: Anwendungen und Erfahrungen mit dem Scientific Use File des Mikrozensus. ZUMA-Arbeitsbericht Nr. 2002/01.
- Schimpl-Neimanns, B./Frenzel, H., 1995: 1-Prozent-Stichprobe der Volks- und Berufszählung. ZUMA-Technischer Bericht Nr. 95/06.
- Seel, B., 2001: Partnerschaftliche Arbeitsteilung und der ökonomische Erklärungsansatz. Zeitschrift für Familienforschung 2: 49-68.
- Statistisches Bundesamt, 1992: Statistisches Jahrbuch 1992 für die Bundesrepublik Deutschland. Stuttgart: Metzler-Poeschel.
- Statistisches Bundesamt (Hrsg.), 2000: Fachserie 1: Bevölkerung und Erwerbstätigkeit. Reihe 1 Gebiet und Bevölkerung. Stuttgart: Metzler-Poeschel.
- Steiner, V./Lauer, C., 2000: Private Erträge von Bildungsinvestitionen in Deutschland. ZEW Discussion Paper 00-18: Mannheim.
- Strohmeier, K.P., 1993: Pluralisierung und Polarisierung der Lebensformen in Deutschland. Aus Politik und Zeitgeschichte 17: 11-22.
- Thornton, A./Axinn, W.G./Teachman, J.D., 1995: The Influence of School Enrollment and Accumulation on Cohabitation and Marriage in Early Adulthood. American Sociological Review 60, October: 762-774.
- Tölke, A., 1995: Geschlechtsspezifische Aspekte der Berufs- und Familienentwicklung. S.489-504 in: Nauck, B./Onnen-Isemann, C. (Hrsg.), Familie im Brennpunkt von Wissenschaft und Forschung. Neuwied: Luchterhand.
- Wirth, H., 1996: Wer heiratet wen? Die Entwicklung der bildungsspezifischen Heiratsmuster in Westdeutschland. Zeitschrift für Soziologie 25, 5: 371-394.
- Wirth, H., 2000: Bildung, Klassenlage und Partnerwahl. Eine empirische Analyse zum Wandel der bildungs- und klassenspezifischen Heiratsbeziehungen. Opladen: Leske+Budrich.
- Ziegler, R., 1985: Bildungsexpansion und Partnerwahl. S. 85-106 in: Hradil, S. (Hrsg.), Sozialstruktur im Umbruch. Festschrift für Karl Martin Bolte. Opladen.: Leske+Budrich.

Anhang

Tabelle A1: Randverteilung der Variablen Bildungsniveau und Alter nach Haushaltsmodus, 1970 und 1997, alte und neue Bundesländer (in %)

	Haushaltsmodus					
	Ehepaarhaushalt			Ledigenhaushalt		
	1970	1997		1970	1997	
	West	West	Ost	West	West	Ost
	Spaltenprozent			Spaltenprozent		
Bildungsabschluss						
Frauen						
HSO	48,9	10,6	0,9	34,4	6,7	1,4
HSM	28,7	32,3	4,0	19,7	19,9	5,1
MRO	3,6	3,1	2,2	5,0	3,0	3,2
MRM	14,8	33,1	74,7	28,6	34,6	71,0
ABIO	0,7	1,5	0,2	0,8	2,4	0,5
ABIM	0,8	8,7	4,7	2,2	14,3	5,5
FH/HS	2,5	10,6	13,3	9,3	19,1	13,2
Insgesamt (n)	65779	32471	8178	7666	12230	2554
Männer						
HSO	19,8	5,4	0,4	29,5	6,5	1,0
HSM	48,3	43,7	5,9	38,3	35,6	7,8
MRO	0,7	1,3	1,6	1,1	1,9	2,6
MRM	21,1	23,0	72,8	18,6	24,8	73,5
ABIO	0,3	1,1	0,3	0,8	2,5	0,5
ABIM	1,1	7,2	3,4	1,0	9,5	4,0
FH/HS	8,6	18,3	15,6	10,7	19,2	10,5
Insgesamt (n)	61954	26896	7053	11929	18190	4431
Alterskohorten						
Frauen						
25-29 Jahre	21,6	13,4	11,1	33,8	42,0	47,5
30-34 Jahre	30,1	27,2	25,6	24,6	30,8	27,4
35-39 Jahre	24,1	30,4	31,8	18,2	17,2	15,9
40-44 Jahre	24,2	29,0	31,5	23,4	10,1	9,3
Insgesamt (n)	65779	32471	8178	7666	12230	2554
Männer						
25-29 Jahre	17,3	8,6	7,8	45,9	35,7	42,9
30-34 Jahre	29,5	25,5	22,7	30,5	33,6	31,3
35-39 Jahre	26,3	32,7	33,0	14,2	19,2	16,2
40-44 Jahre	26,8	33,2	36,5	9,4	11,5	9,5
Insgesamt (n)	61954	26896	7053	11929	18190	4431

Quelle: 1%-Stichprobe Volkszählung 1970 und anonymisierte 70%-Unterstichprobe des Mikrozensus 1997 (jeweils ZUMA-Files)

Tabelle A2: Geschätzte Koeffizienten des logistischen Regressions-modells B0 (Übersicht 2), Zeitvergleich Westdeutschland 1970-1997; (Abhängige Variable: Ledigenhaushalt versus Ehepaarhaushalt)

Zeitvergleich Westdeutschland 1970-1997	
Modell B0 ; Übersicht 2	
Referenzkategorie: Verheiratete Frauen mit Fachhochschulabschluss/Hochschulabschluss (FH/HS) im Alter von 40-44 Jahren zum Erhebungszeitpunkt 1970	Beta
Konstante (b ₀)	-1,35
Bildungsabschluss	
Hauptschulabschluss ohne berufliche Ausbildung (HSO)	-1,19*
Hauptschulabschluss mit beruflicher Ausbildung (HSM)	-1,32*
Mittlere Reife ohne berufliche Ausbildung (MRO)	-0,71*
Mittlere Reife mit beruflicher Ausbildung (MRM)	-0,50*
Abitur ohne berufliche Ausbildung (ABIO)	-0,56*
Abitur mit beruflicher Ausbildung (ABIM)	-0,19*
Alterskohorte	
25-29 Jahre	0,98*
30-34 Jahre	0,20*
35-39 Jahre	-0,11
Geschlecht	
Männer	-1,03*
Beobachtungszeitpunkt	
1997	0,31*
Interaktion Bildung mit Geschlecht	
HSO*Männer	1,61*
HSM*Männer	0,90*
MRO*Männer	0,92*
MRM*Männer	0,32*
ABIO*Männer	1,13*
ABIM*Männer	0,34*
Interaktion Bildung mit Alter	
HSO*25-29 Jahre	-0,40*
HSO*30-34 Jahre	-0,12*
HSO*35-39 Jahre	-0,02

Fortsetzung Tabelle A2

HSM*25-29 Jahre	-0,27*
HSM*30-34 Jahre	-0,14*
HSM*35-39 Jahre	-0,02
MRO*25-29 Jahre	-0,23*
MRO*30-34 Jahre	-0,22*
MRO*35-39 Jahre	-0,08
MRM*25-29 Jahre	-0,25*
MRM*30-34 Jahre	-0,19*
MRM*35-39 Jahre	-0,10
ABIO*25-29 Jahre	-0,27
ABIO*30-34 Jahre	-0,32*
ABIO*35-39 Jahre	-0,16
ABIM*25-29 Jahre	-0,19*
ABIM*30-34 Jahre	-0,24*
ABIM*35-39 Jahre	-0,06
Interaktion Bildung mit Beobachtungszeitpunkt	
HSO*1970	-0,06
HSM*1970	0,12*
MRO*1970	-0,06
MRM*1970	-0,14*
ABIO*1970	0,32*
ABIM*1970	-0,07
Interaktion Alterskohorte mit Geschlecht	
Männer*25-29 Jahre	1,08*
Männer*30-34 Jahre	0,76*
Männer*35-39 Jahre	0,41*
Interaktion Alterskohorte mit Beobachtungszeitpunkt	
25-29 Jahre*1970	1,06*
30-34 Jahre*1970	0,75*
35-39 Jahre*1970	0,41*
Interaktion Beobachtungszeitpunkt mit Geschlecht	
1970 *Männer	0,58*
Pseudo-R ²	0,161
Fallzahl	237.115

*p < 0,05

Quelle: 1%-Stichprobe Volkszählung 1970 und anonymisierte 70%-Unterstichprobe des Mikrozensus 1997 (jeweils ZUMA-Files)

Tabelle A3: Geschätzte Koeffizienten des logistischen Regressionsmodells B0 (Übersicht 3), Ost-Westvergleich 1997; (Abhängige Variable: Ledigenhaushalt versus Ehepaarhaushalt)

Ost-Westvergleich 1997	
Modell B0 ; Übersicht 3	
Referenzkategorie: Verheiratete Frauen mit Fachhochschulabschluss/Hochschulabschluss (FH/HS) im Alter von 40-44 Jahren in Westdeutschland	Beta
Konstante (b ₀)	-1,36
Bildungsabschluss	
Hauptschulabschluss ohne berufliche Ausbildung (HSO)	-0,92*
Hauptschulabschluss mit beruflicher Ausbildung (HSM)	-1,03*
Mittlere Reife ohne berufliche Ausbildung (MRO)	-0,63*
Mittlere Reife mit beruflicher Ausbildung (MRM)	-0,72*
Abitur ohne berufliche Ausbildung (ABIO)	-0,07
Abitur mit beruflicher Ausbildung (ABIM)	-0,27*
Alterskohorte	
25-29 Jahre	2,44*
30-34 Jahre	1,32*
35-39 Jahre	0,53*
Geschlecht	
Männer	0,03
Ost-West-Variable	
Ost	-0,99*
Interaktion Bildung mit Geschlecht	
HSO*Männer	1,07*
HSM*Männer	0,72*
MRO*Männer	0,78*
MRM*Männer	0,52*
ABIO*Männer	0,87*
ABIM*Männer	0,40*
Interaktion Bildung mit Alterskohorte	
HSO*25-29 Jahre	-0,65*
HSO*30-34 Jahre	-0,12
HSO*35-39 Jahre	-0,02

Fortsetzung Tabelle A3

HSM*25-29 Jahre	-0,28*
HSM*30-34 Jahre	-0,23*
HSM*35-39 Jahre	-0,10
MRO*25-29 Jahre	-0,32*
MRO*30-34 Jahre	-0,28
MRO*35-39 Jahre	-0,05
MRM*25-29 Jahre	-0,21*
MRM*30-34 Jahre	-0,25*
MRM*35-39 Jahre	-0,11
ABIO*25-29 Jahre	-0,37*
ABIO*30-34 Jahre	-0,40*
ABIO*35-39 Jahre	-0,17
ABIM*25-29 Jahre	-0,19
ABIM*30-34 Jahre	-0,30*
ABIM*35-39 Jahre	-0,08
Interaktion Bildung mit Ost-West	
HSM* Ost	1,60*
HSM* Ost	1,29*
MRO* Ost	0,84*
MRM* Ost	0,53*
ABIO* Ost	0,25
ABIM* Ost	0,29*
Interaktion Alterskohorte mit Geschlecht	
Männer*25-29 Jahre	0,32*
Männer*30-34 Jahre	0,22*
Männer*35-39 Jahre	0,08
Interaktion Alterskohorte mit Ost-West	
25-29 Jahre* Ost	0,50*
30-34 Jahre* Ost	0,31*
35-39 Jahre* Ost	0,12*
Interaktion Ost-West mit Geschlecht	
Ost *Männer	0,15*
Pseudo-R ²	0,150
Fallzahl	112.003

*p < 0,05

Quelle: 1%-Stichprobe Volkszählung 1970 und anonymisierte 70%-Unterstichprobe des Mikrozensus 1997 (jeweils ZUMA-Files)

BERUFSSKLASSIFIKATION UND MESSUNG DES BERUFLICHEN STATUS/PRESTIGE

JÜRGEN H.P. HOFFMEYER-ZLOTNIK & ALFONS J. GEIS

In den ZUMA-Nachrichten 47 und 48 haben wir uns zum „Stand der Berufsvercodung“ und zur „Kompatibilität von ISCO-68, ISCO-88 und KldB-92“, den drei für die Klassifikation von Berufen in Deutschland gängigsten Codierungen, geäußert. Als weiterer Beitrag zur Berufsklassifikation soll hier dargestellt werden, welche Möglichkeiten eine Berufsklassifikation nach ISCO als demographisches Hintergrundmerkmal bietet. Hierunter sind die Umsetzungen der ISCO-Codierung in Skalen des sozio-ökonomischen Status, des beruflichen Prestige und von Klassenlage zu verstehen. Des weiteren wird auf nationale Alternativen zu Status und Prestige, die auf der „Stellung im Beruf“ basieren, kurz hingewiesen werden.

1. Sozio-ökonomischer Status, berufliches Prestige und Klassenlage

„Status“ definiert die Position eines Menschen in Bezug auf die Positionen anderer. Damit positioniert „Status“ die Stellung von Menschen in hierarchischen Zusammenhängen. Präzisiert auf „sozialen Status“ wird über diesen eine Person in der Hierarchie der Gesellschaft, in der er sich bewegt, verortet. Dieses geschieht im allgemeinen über die Indikatoren „Bildung“, „Beruf“ und „Einkommen“. Hiermit wird der „soziale Status“ zum „sozio-ökonomischen Status“ (Duncan 1961). Betrachtet man jedoch Mobilität, so konzentriert sich diese empirisch sehr schnell auf eine nach ihrem Prestige geordnete Liste von Berufen (obwohl soziale Mobilität viel mehr umfasst als berufliche Mobilität über die Zeit). Damit wird hierarchischer „Status“, also „Prestige-Status“, zu einer Sonderform des „Status“ im Sinne von „Position“. Gemeint ist die Position, die Beziehungen zu den jeweils anderen Schichtpositionen etabliert.

Unter „Prestige“ wird zunächst das Ansehen verstanden, das einer Stellung anhaftet. „Berufs-Prestige“ ist also das Ansehen, das einer beruflichen Tätigkeit und einer beruflichen Position oder Stellung zugeschrieben wird. Damit wird „Berufs-Prestige“ in den

modernen industriellen bis postindustriellen Gesellschaften zur wichtigsten Komponente des allgemeinen Ansehens einer Person, sofern man davon ausgehen kann, dass soziale Mobilität in einer halbwegs „offenen“ Gesellschaft möglich ist. Die soziale Mobilität in einer „offenen“ Gesellschaft wird empirisch erfassbar über berufliche Mobilität.

Die „Klassenlage“ verbindet Informationen über Beschäftigten-Status und berufliche Tätigkeit. Damit ist die Klassenlage im Gegensatz zu Prestige und sozio-ökonomischem Status kein hierarchisches sondern ein kategoriales Differenzierungsschema, das in die drei Obergruppen „Arbeitgeber“, die „Selbständigen“ und die „Arbeitnehmer“ unterscheidet. Die Konzeption der Klassenlage ist als Resultat von Marktlage und Arbeitssituation, in der sich eine beschäftigte Person befindet, zu sehen (Erikson/Goldthorpe/Portocarero 1979; Erikson/Goldthorpe 1992).

2. „Beruf“ als Status-Variable

Bei der Definition von „Beruf“ gibt es nicht nur Unterschiede von Staat zu Staat bzw. von Kultur zu Kultur, denn durch eine unterschiedliche Organisation der Arbeit können „nominell identische Berufe nicht notwendig dieselben Tätigkeiten umfassen“. Auch können „Berufe ... immer spezifischer definiert werden“ (vgl. Treiman 1979: 126f.). Wichtig ist es, dass eine Skala benutzt wird, die die nationalen Berufsstatus-Hierarchien angemessen repräsentiert und die richtig standardisiert ist. Hierzu dient die Standardklassifikation der Berufe der „International Standard Classification of Occupations“ (ISCO).

Die Wichtigkeit der Variable „Beruf“ wird dadurch unterstrichen, dass die beiden anderen Status-Variablen (Bildung und Einkommen) an separatem Gewicht verloren haben und sich heute fast ausschließlich in der Variable „Beruf“ wiederfinden. Die „Bildung“ ist über länderspezifische Bildungssysteme (allein in der Bundesrepublik kann man den Basisabschluss, bundeslandbedingt, alternativ nach 9 oder 10 Schuljahren und die Hochschulreife nach 12 oder 13 Schuljahren erreichen), eine in vielen Ländern allgemein zu verzeichnende Bildungsexpansion in den letzten 20 bis 30 Jahren, sowie ein Überangebot an Ausbildungsnachfragern als Status-Variable nicht mehr tauglich. Das Gefüge von „Bildung“ zu „Beruf“ ist nachhaltig beeinträchtigt. Höhere Bildung ist kein Status-Plus mehr, sondern bietet lediglich eine bessere Ausgangssituation im Verteilungskampf um das knapper werdende Gut „Arbeit“. Ebenfalls finden reale Einkommensverluste statt, bedingt durch eine lang anhaltende Stagnation bei den Lohn- und Gehaltssteigerungen. Dieses dokumentiert, sofern das Einkommen zu hoch gewichtet würde, eine negative soziale Mobilität. Außerdem weist die Einkommensmessung Messprobleme auf, die es geraten erscheinen lassen, „Einkommen“ bei der Bestimmung des sozio-ökonomischen Status nicht zu hoch zu bewerten (siehe Hoffmeyer-Zlotnik/Warner 1998). Allerdings

steht die berufliche Tätigkeit auch mit der Variable „Einkommen“ in einem engen Verhältnis: Je höher das Prestige eines Berufes/einer Tätigkeit ist, desto höher ist auch das durch die Ausübung dieser Tätigkeit erzielte Einkommen. Dieses bedeutet allerdings nicht, dass jede Tätigkeit mit hohem Prestige auch ein hohes Einkommen bedingt, und es bedeutet auch nicht, dass identische Prestige-Werte auf unterschiedlichen Positionen identische Einkommenshöhen garantieren.

Der „Beruf“ als sozialstatistisches Hintergrundmerkmal basiert auf der impliziten Annahme, die Verortung einer Person im sozialen Gefüge einer Gesellschaft sei vor allem durch die in einer arbeitsteiligen Gesellschaft spezifische Erwerbstätigkeit gegeben (Mayer 1979: 81). Bildung und Einkommen stehen mit der Variable der beruflichen Position in engem Zusammenhang. Die Wichtigkeit von beruflicher Tätigkeit und Position werden allerdings damit begründet, dass diese als wichtige Faktoren für Lebensstile, Verhaltensweisen, Einstellungen und Gruppenhandeln gesehen werden (Mayer 1979: 81).

3. Prestige- und Status-Skalen, basierend auf ISCO-68 und ISCO-88

Um möglichst vielen Konzepten gerecht zu werden, ist es sinnvoll, die Variable „Beruf“ so detailliert zu erfassen, dass eine Vercodung nach unterschiedlichen Forschungsinteressen und nach der internationalen Standardklassifikation, dem ISCO, möglich ist. Zwar sind einige Prestige-Skalen sowie die Klassenlage auch über eine KldB-92-Klassifikation generierbar; allerdings führt die KldB-92-Klassifikation zu einer spezifisch deutschen Klassifikation. Hierbei ist die KldB-92-Klassifikation in der Codierung auch nicht viel weniger aufwendig als eine ISCO-Klassifikation.

Im folgenden sollen die wichtigsten Skalen zu Prestige und sozio-ökonomischem Status dargestellt werden. Alle vorgestellten Skalen benötigen als Ausgangsdaten eine ISCO-Vercodung nach ISCO-68 oder ISCO-88. Zunächst auf der Basis von ISCO-68 entwickelt, sind alle für den internationalen Gebrauch erstellten Skalen heute auch auf ISCO-88 umgestellt worden. Nur die als nationale Skala für Deutschland erstellte Magnitude-Prestige-Skala von Wegener ist nicht auf ISCO-88, sondern statt dessen auf die nationale deutsche Klassifikation der KldB-92 umgestellt worden (Frietsch/Wirth 2001).

Einen Vergleich der unterschiedlichen Prestige- und Status-Skalen bieten Wolf 1995 und Ganzeboom/Treiman 2003.

3.1 Die internationale Berufsprestige-Skala von Treiman (SIOPS)

Um ein Instrument für den internationalen Vergleich zur Verfügung zu haben, wurde in den 70-er Jahren der Versuch unternommen, eine standardisierte berufliche Prestige-Skala zu entwickeln (Treiman 1975; 1977). Bei dieser standardisierten Prestige-Skala erhält jeder bekannte Beruf in jeder Gesellschaft denselben Wert. Die Grundvoraussetzung ist die Annahme des zeitlich und räumlich invarianten Charakters beruflicher Prestige-Hierarchien (Treiman 1977). Durch die standardisierte Prestige-Skala ist eine gültige Schätzung der beruflichen Prestige-Hierarchie in jedem Land möglich.

Erstellt wurde die erste international messende Prestige-Skala dadurch, dass in 55 Ländern, von der spätindustriellen Gesellschaft bis zur agrarischen Gesellschaft, in Surveys die Befragungspersonen die Aufgabe erhalten haben, eine Menge von Berufsbezeichnungen (basierend auf einer Modifikation der „unit groups“ der ISCO-68: insgesamt 509 Berufsbezeichnungen) hinsichtlich ihres sozialen Ansehens zu beurteilen und zu rangreihen (zur Skalenbildung siehe Treiman 1977, Kap. 8 und 9). Der Grad der internationalen Übereinstimmung wird bestimmt über eine Korrelation zwischen den Prestige-Werten identischer Berufe in unterschiedlichen Ländern. Treiman (1975: 197) berichtet von durchschnittlichen Korrelationen zwischen Paaren von 55 Ländern von $r = 0.81$. Er kommentiert, dass idiosynkratische Prestige-Beurteilungen zumeist in Studien von relativ schlechter Qualität auftreten würden. Er vermutet daher eine durchschnittliche Korrelation von 0.8 als unteren Wert. Um die Güte der Skala für das einzelne Land aber zu prüfen, wurde die Korrelation zwischen der Prestige-Skala eines jeden einzelnen Landes und einer Version der Standard-Skala, in der das betreffende Land gestrichen war, gerechnet. Die durchschnittliche Korrelation zwischen der so geänderten Standardskala mit den ursprünglichen landesspezifischen Prestige-Werten eines jeden der 55 Länder betrug 0.89. Nur 7 Korrelationen waren kleiner als 0.8 und 15 fielen in den Bereich zwischen 0.8 und 0.9 (Treiman 1975: 198f.).

Bei Treimans „Standard International Occupational Prestige Scale“ (SIOPS-Skala) war der empirische Befund aus Ländern von der Agrargesellschaft bis zur spätindustriellen Gesellschaft das Ausgangsmaterial. Daher sei es für das Bewertungssystem nach Treiman unproblematisch, sofern es differenziert genug angelegt ist, ob man mit dieser so konstruierten Skala eine frühindustrielle oder eine postindustrielle Gesellschaft untersucht. Fällt die Orientierung am Markt jedoch in einem geänderten Schichtungssystem weg, und treten andere Bewertungskriterien statt dessen an diese Stelle, wie es in einer sozialistischen Gesellschaft der Fall ist, so ändert sich auch die Bewertung und die Hierarchie der Berufe. Damit verliert eine Prestige-Skala, wie die von Treiman konstruierte, ihre das gesellschaftliche Schichtungssystem beschreibende Gültigkeit: Bei einer gesellschaftlich bedingten Aufwertung der Produktion von Gütern und Abwertung

von Dienstleistungen entsteht ein abweichendes Bewertungssystem von Berufstiteln (vgl. Geis/Hoffmeyer-Zlotnik 1991). Eine entsprechende Abweichung bei der Nutzung einer international vergleichenden Prestige-Skala ist z.B. auch möglich in traditionellen Agrarstaaten, in denen das Schichtungssystem der Industriegesellschaft nicht gilt. Bezogen auf das Europa der EU und assoziierte Länder gilt seit Beginn der 1990-er Jahre, dass die gesellschaftliche Schichtung nach dem Wegfallen des Sozialismus auf vergleichbaren Bewertungskriterien aufbaut. Damit ist die Anwendung einer ISCO-basierten, d.h. eine auf der beruflichen Tätigkeit aufbauende Prestige-Skala, gerechtfertigt.

Die internationale Berufsprestige-Skala von Treiman (SIOPS) wurde in den 90er Jahren auf die ISCO-88 Codierung umgestellt. Soweit möglich, wurden auf der Basis der 390 „unit-groups“ die vorhandenen eins-zu-eins Übertragungen von ISCO-68 zu ISCO-88 aus dem Handbuch übernommen. Dort, wo eine eins-zu-eins Übertragung nicht möglich war, wurde ein mittlerer Wert aller zu einer „unit-group“ zählenden Berufe (in der Einteilung gemäß Treiman 1977) ermittelt und den ISCO-88-Werten in gleicher Logik zugeordnet, wie dieses zuvor schon mit ISCO-68 geschehen war. Entsprachen einer „unit-group“ keine Einträge, dann wurde die Wertvergabe an ähnlichen „unit-groups“, für die ein Wert erstellt werden konnte, orientiert (Ganzeboom et al. 1996).

3.2 Die internationale Skala des sozio-ökonomischen Status von Ganzeboom et al. (ISEI)

Neben der Prestige-Skala von Treiman ist 1992 eine weitere, international messende Bewertungs-Skala von gesellschaftlicher Ordnungsstruktur publiziert worden: der „Standard International Socio-Economic Index of Occupational Status“ (ISEI-Skala) von Ganzeboom, De Graaf, Treiman und de Leeuw (1992). Hierbei geht es nicht um berufliches Prestige, sondern um den sozio-ökonomischen Status: Die berufliche Tätigkeit wird mit Informationen über das Einkommen und die Bildung kombiniert. Ausgangsmaterial waren die Informationen über Bildung, Beruf und Einkommen von etwa 74.000 vollzeitbeschäftigten männlichen Befragungspersonen im Alter zwischen 21 und 64 Jahren (Ganzeboom et al. 1992: 13f.). Die Daten stammen aus insgesamt 31 Untersuchungen aus 16 verschiedenen Ländern.

Hinter dieser Skala steht nicht mehr die Überlegung, dass jede berufliche Tätigkeit ein Ansehen hat, das bewertbar und zu anderen beruflichen Tätigkeiten über eine Rangreihung in Beziehung zu setzen ist. Die Überlegung zu dieser Skala geht davon aus, dass jede berufliche Tätigkeit einen bestimmten Bildungsgrad erfordert und durch ein bestimmtes Lohnniveau entlohnt wird. Technisch ist der Index an die „unit groups“ der beruflichen Tätigkeiten der ISCO-68 Klassifikationen geknüpft, die, entsprechend den

Bedingungen von erforderlichem Bildungsgrad und Höhe des Arbeitseinkommens, neu skaliert werden. Bis zu einem gewissen Grad folgt dieser Index damit den Überlegungen, die auch hinter der ISCO-Revision von 1988 stehen. Allerdings sind die „skill levels“ der ISCO-88 eher Paten als Merkmale der Skalenkonstruktion und die „skill specialisations“ stehen außerhalb dieses Konzeptes. Damit bleibt die Konstruktionsgrundlage die berufliche Tätigkeit und das Ausgangsinstrument die ISCO-68.

1996 wurde dieser Index auf die Datenbasis und die Logik der ISCO-88 übertragen (Ganzeboom et al. 1996). Neu hierbei ist teilweise die Organisation spezieller Berufsgruppen durch Hinzunahme oder Auslassung bisheriger Kategorien, eine Differenzierung in den Kategorien der leitenden Positionen, sowie die Berücksichtigung des „skill-level“. Ein Großteil der Werte sind jedoch mit der auf ISCO-68 aufbauenden ISEI-Skala identisch.

3.3 Die Messung nominaler Klassenkategorien von Erikson, Goldthorpe und Portocarero (EGP)

Die nominale Typologie zur Kombination des Berufes mit Informationen zur beruflichen Stellung wurde ursprünglich für speziell britische Untersuchungen entworfen und von Erikson, Goldthorpe und Portocarero (1979) in einer Studie über drei Länder zu einem Standard für internationale Vergleiche weiterentwickelt. Die heute vorliegende Klassifikation beruht auf der Kooperation von Erikson und Goldthorpe (1992) mit dem CASMIN-Projekt (beschrieben in den ZUMA-Nachrichten 46 von Brauns/Steinmann/Haun, 2000). Die für die Konstruktion der „Klassenlage“ benötigten Variablen sind der „ausgeübte Beruf“ (für den internationalen Vergleich nach ISCO zu klassifizieren) und die „Stellung im Beruf“ inklusive einer Differenzierung nach Selbständigen, abhängig Beschäftigten und mithelfenden Familienangehörigen.

Da das Verfahren von Erikson/Goldthorpe/Portocarero nicht ausreichend dokumentiert war, um dieses auf andere Datensätze übertragen zu können, haben Ganzeboom et al. (1992) ein Modul erstellt, das es seither ermöglicht, die EGP-Kategorien auf der Basis von ISCO-68 herzuleiten. Deren Vorgehen bestand darin, zuerst die Berufsbezeichnung als Hauptinformation, dann die berufliche Stellung als Korrekturhilfe zu nutzen und in ein einfaches Schema, welches die Variablen Selbständigkeit und Stellung im Beruf kombiniert, einzupassen.

Die in Tabelle 1 wiedergegebene EGP-Recodierung wurde von Ganzeboom et al. in Anlehnung an die Dokumentation von Erikson und Goldthorpe im CASMIN-Projekt vorgenommen.

Tabelle 1: Die Kategorien der EGP-Klassen

	<i>EGP 11</i>	<i>Description</i>
I	1	Higher Managerial and Professional Workers
II	2	Lower Managerial and Professional Workers
IIIa	3	Routine Clerical Work
IIIb	4	Routine Service and Sales Work
IVa	5	Small Self-Employed with Employees
IVb	6	Small Self-Employed without Employees
V	7	Manual Supervisors
VI	8	Skilled Manual Workers
VIIa	9	Semi- and Unskilled Manual Workers
VIIb	10	Agricultural Labour
IVc	11	Self-Employed Farmers

Ganzeboom/Treiman 2003: 172

Auch die EGP-Skala wurde Mitte der 90er Jahre von Ganzeboom et al. (1996) auf der Basis der ISCO-88 modernisiert.

3.4 Die Magnitude-Prestige-Skala (MPS) von Wegener

Neben den für den internationalen Vergleich entwickelten Skalen zu Prestige, Status und Klassenlage gibt es auch nationale Skalen. Die für die Sozialforschung der Bundesrepublik wichtigste nationale Prestige-Skala, die auf der ISCO-68 aufbaut, ist die **Magnitude-Prestige-Skala (MPS)** von Wegener (1988). Diese Skala wurde aus Daten von zwei Studien entwickelt, die 1979 und 1980 in den alten Ländern der Bundesrepublik Deutschland erhoben wurden. Der Ausgangspunkt war eine Skalierung von 50 verschiedenen Berufen hinsichtlich ihres gesellschaftlichen Ansehens mit dem Verfahren der Magnitude-Skalierung. Um diese 50 von den Befragten eingeordneten Berufe wurden die weiteren Berufe, unter anderem unter Rückgriff auf die Prestige-Skala von Treiman (1977), eingeordnet. Die Magnitude-Prestige-Skala von Wegener ist bisher nicht auf ISCO-88 umgestellt, worden.

4. Probleme bei der Codierung für unzureichend beschriebene Berufsangaben

Berufsangaben, die ungenügende Informationen enthalten, so dass keine eindeutige Klassifikation nach den Regeln der ISCO möglich ist, können oft dennoch über Sondercodes zu Prestige- oder Status-Werten zugeordnet werden. SIOPS, ISEI und auch MPS erfassen die unvollständigen Codes, indem bei der ISCO-68 der drei- statt der vierstellige und bei

der ISCO-88 der zwei statt der dreistellige Code die Ausgangsbasis für die Zuordnung von Prestige- oder Status-Wert bildet. Statt eines spezifisch den Beruf bezeichnenden Codes wird in solchen Fällen entweder ein Skalenwert relational zur Variationsbreite des Prestige- oder Status-Wertes errechnet oder es wird der Durchschnittswert für die Obergruppe ermittelt und als Sondercode herangezogen. Die Ermittlung eines Durchschnittswertes innerhalb einer ISCO-Klassifikation kann dann problematisch werden, wenn innerhalb der für die Klassifikation heranzuziehenden einzelnen Berufe einige mit hohem oder niedrigem Prestige- oder Status-Wert selten vorkommen. In diesem Fall kann die Mittelwert-Bildung verzerrt sein.

Als weiterer Sonderfall muss angeführt werden, dass es auch nationale Codes bei der Klassifikation nach ISCO gibt. Des weiteren gibt es für unvollständige Angaben auch spezifisch vom ZUMA entwickelte Codes, damit eine Vercodung nach ISCO möglich bleibt. Auch für diese von der internationalen Klassifikation abweichenden Codes wurde vom ZUMA eine Zuordnung von Prestige- oder Status-Werten ermittelt.

5. Probleme bei der Validität von Prestige-Werten

Prestige-Skalen stellen hierarchische Rangordnungen von gesellschaftlichen Bewertungsstrukturen dar. Hierbei kommt es einerseits darauf an, wie diese Rangordnungen in einer spezifischen Gesellschaft verankert sind und andererseits ist ausschlaggebend, worüber Prestige definiert ist. Bei Treiman (1977) sind die Attribute der beruflichen Position selbst Kriterium für die Rangordnung. Die Skala stellt damit eine Bewertungsstruktur dar, die die Wahrnehmung des zu einer Berufsausübung notwendigen Aufwands an Bildung und Ausbildung berücksichtigt und die sich an der Komplexität der Tätigkeit und der Autonomie des Handelnden orientiert. So gesehen ist es gleich, ob man nach diesen Kriterien in einer industriellen oder einer postindustriellen Gesellschaft eine Hierarchie der Berufe herstellt. Problematisch wird es dann, wenn die gewohnten Kriterien für die Hierarchiebildung nicht mehr stimmen, z.B. in einer Agrargesellschaft oder in einer sozialistischen Gesellschaft. Sowohl in einer Agrargesellschaft als auch in einer sozialistischen Gesellschaft gelten andere Bewertungssysteme für berufliche Tätigkeiten, als jene für eine industrielle oder postindustrielle, am Markt orientierte Gesellschaft (siehe auch Geis/Hoffmeyer-Zlotnik 1991).

6. Nationale Alternativen zu den ISCO-basierten Skalen

Alternativen zu den an der Berufsangabe orientierten Prestige- oder Status-Skalen stellen Skalen dar, die auf der „Stellung im Beruf“ aufbauen. Hiervon gibt es eine größere Anzahl von nationalen Skalen für die Bundesrepublik Deutschland, die bei Wolf (1995) kurz dargestellt sind. Die in der Konstruktion einfachste der auf der „Stellung im Beruf“

aufbauenden Skalen soll hier als eine Alternative zu den ISCO-basierten Skalen dargestellt werden: die Skala „Autonomie des beruflichen Handelns“ von Hoffmeyer-Zlotnik (1998), die berufliches Prestige misst.

Aufbauend auf der „Stellung im Beruf“, die in der Bundesrepublik Deutschland über die „Demographischen Standards“ sehr tief gegliedert abgefragt wird (siehe: Statistisches Bundesamt 1999), wird eine Skala zur „Autonomie des beruflichen Handelns“ entwickelt, die die erwerbstätigen Befragten in 5 Kategorien beruflichen Prestige zusammenfasst. Diese für Umfragen in der Bundesrepublik Deutschland erstellte Prestige-Skala stellt dann eine Alternative zu den ISCO-basierten Status- und Prestige-Skalen dar, wenn eine Differenzierung der Befragten in wenigen Kategorien für die Analyse ausreichend ist.

Die Grundlage für die Skala „Autonomie des Handelns“ ist die tiefe Klassifikation der „beruflichen Stellung“, die über die zur Statusbestimmung irrelevanten versicherungstechnischen Kategorien („Selbständiger“, „Beamter“, „Angestellter“, „Arbeiter“) hinausgehend, pro Gruppe eine Differenzierung der Stellung entsprechend unterschiedlicher Tätigkeitsmerkmale aufweist: Die Selbständigen, gegliedert in die Gruppen „selbständige Landwirte“, „akademische freie Berufe“ und „Selbständige in Handel, Gewerbe, Industrie, Dienstleistungen“, werden unterteilt nach der Größe des Betriebes (Landwirte nach ha-Größe der Nutzfläche, die beiden anderen Gruppen nach der Anzahl der Mitarbeiter); bei den Beamten wird nach dem Dienstrecht (einfacher, mittlerer, gehobener, höherer Dienst) unterschieden, über das sich auch die Tätigkeit und die damit verbundene Autonomie des Handelns definiert; die Arbeiter werden entsprechend ihrer Ausbildung und damit nach einer Hierarchie ihrer Einsetzbarkeit und Verantwortlichkeit für die Tätigkeit in „Ungelernte“, „Angelernte“, „Facharbeiter“ und „Meister/Poliere“ untergliedert; die Angestellten werden nach der Differenziertheit der Tätigkeit und nach der damit verbundenen Verantwortung klassifiziert in diejenigen, die a) eine „einfache Tätigkeit“ haben, b) „schwierige Aufgaben nach allgemeiner Anweisung selbständig erledigen“, c) „selbständige Leistungen in verantwortungsvoller Tätigkeit erbringen oder begrenzte Verantwortung für die Tätigkeit anderer tragen“ und schließlich d) die „umfassende Führungsaufgaben und Entscheidungsbefugnisse“ innehaben.

Betrachtet man die drei Gruppen der abhängig Beschäftigten (Beamte, Arbeiter, Angestellte), so ergibt sich für jede Gruppe eine hierarchische Unterteilung nach unterschiedlichen Stufen der Autonomie des Handelns, beginnend mit niedriger Handlungsautonomie bei einfacher Tätigkeit, bis hin zu hoher Handlungsautonomie bei der Ausübung umfassender Führungsaufgaben. In einem zweiten Schritt ist diese Skala der Handlungsautonomie der abhängig Beschäftigten zu erweitern um die Gruppen der Selbständigen, deren Handlungsautonomie sich in Produktion und Dienstleistung sowie im akademisch

freien Bereich in der Regel aus der Größe des Betriebes ergibt, auch wenn die jeweils unterste Ebene dieser Gruppen auf unterschiedlichem Niveau liegt. Im Bereich der Landwirtschaft ist die Betriebsgröße ein zentrales, wenn auch nicht das einzige Kriterium für die Autonomie des Handelns. In einem dritten Schritt werden die „mithelfenden Familienangehörigen“ mit einem mittleren Wert berücksichtigt, da diese Gruppe sehr heterogen ist und deren Verantwortung für die Tätigkeit ein hohes Niveau erreichen kann.

Tabelle 2: „Stellung im Beruf“ nach Autonomie des Handelns

Wert „Autonomie“	Code „Stellung im Beruf“
1 - niedrig	60, 61
2	40, 51, 62, 10, 11, 12, 13
3	41, 52, 63, 15, 21, 30
4	42, 50, 53, 64, 16, 22
5 - hoch	43, 54, 17, 23, 24

Hoffmeyer-Zlotnik 1998: 61

Zur Überprüfung der Güte der Skala „Autonomie des Handelns“ wird die internationale Prestige-Skala von Treiman herangezogen. Wolf (1995) hat einen weit umfangreicheren Vergleich über eine größere Auswahl sowohl ISCO-basierter als auch Stellung-im-Beruf-basierter Indices vorgenommen.

Erste Voraussetzung für eine Reliabilitätsprüfung der „Autonomie des Handelns“ im Vergleich zur Treiman-Prestige-Skala ist eine Transformation der Treiman-Skala, die in der Bundesrepublik Deutschland einen Wertebereich von 14 bis 78 aufweist, auf eine 5-er Skala von „1 = niedrig“ bis „5 = hoch“, damit beide Skalen den gleichen Wertebereich abbilden. Die Schnittstellen sind nach Plausibilität festgesetzt worden: Mit dem Wert 32 als Schnittstelle sind vor allem niedrig bewertete manuelle Tätigkeiten mit geringer Handlungsautonomie in der Gruppe 1 konzentriert. Die Gruppe 2 mit den Wertebereichen 33 bis 41 umfasst abhängige Beschäftigung in der Produktion und im Dienstleistungsbereich, die eine geringe Spezialisierung voraussetzen. Die Gruppe 3 mit den Wertebereichen 42 bis 50 umfasst Tätigkeiten, die einen mittleren Ausbildungsabschluss voraussetzen und die in einem beschränkten Maße verantwortungsvolles Handeln erfordern. Die Gruppe 4 mit den Wertebereichen 51 bis 63 umfasst Tätigkeiten in abhängiger oder selbständiger Position, die eine Fachhochschul- oder Hochschulausbildung voraussetzen aber kein sehr hohes Prestige haben. Mit dem Wert 64 beginnen die leitenden Tätigkeiten und die freiberuflichen Akademiker der Gruppe 5, die ein hohes Ansehen und/oder ein größeres Büro haben. Eine Korrelation zwischen „Autonomie des beruflichen Handelns“ von Hoffmeyer-Zlotnik (1998) und dem

„Berufs-Prestige“ von Treiman (1979) beträgt $r = .7860$.

7. Fazit

Bei Interesse an einer tiefer differenzierenden Hintergrundvariable wie „berufliches Prestige“ oder „sozio-ökonomischer Status“, oder bei der Analyse nach der „Klassenlage“ ist das Erfassen der Berufe der zu Analysierenden und eine Klassifikation der Berufe die Voraussetzung. Hierzu muss eines der Instrumente zur Klassifikation der Berufe benutzt werden. Im internationalen Vergleich ist eine Beschränkung auf die ISCO-Codierung nach ISCO-68 oder ISCO-88 zwingend, bei nationalen Datensätzen können auch die nationalen Klassifikationssysteme benutzt werden - in der Bundesrepublik Deutschland die KldB-92 des statistischen Bundesamtes. Reicht als Hintergrundvariable eine grobe Differenzierung von beruflichem Prestige, so empfiehlt sich, nicht die aufwendige Berufsklassifikation nach ISCO oder KldB vorzunehmen, sondern eine Skala, die auf der „Stellung im Beruf“ aufbaut.

Korrespondenzadresse

PD Dr. Jürgen H.P. Hoffmeyer-Zlotnik
ZUMA
Postfach 12 21 55
68 072 Mannheim
email: hoffmeyer-zlotnik@zuma-mannheim.de

Literatur

- Brauns, H./Steinmann, S./Haun, D., 2000: Die Klassifikation des Klassenschemas nach Erikson, Goldthorpe und Portocarero (EGP) am Beispiel nationaler Datenquellen aus Deutschland, Großbritannien und Frankreich. ZUMA-Nachrichten 46: 7-63.
- Duncan, O.D., 1961: A Socioeconomic Index for all Occupations. S. 109-138 in: Reiss, A.J. jr. (ed.), Occupations and Social Status. New York: Free Press of Glencoe.
- Erikson, R./Goldthorpe, J.H., 1992: The Constant Flux. A Study of Class Mobility in Industrial Societies. Oxford: Clarendon Press.
- Erikson, R./Goldthorpe, J.H./Portocarero, L., 1979: Intergenerational Class Mobility in Three Western European Societies. British Journal of Sociology 30: 415-441.
- Frietsch, R./Wirth, H., 2001: Die Übertragung der Magnitude-Prestigeskala von Wegener auf die Klassifizierung der Berufe. ZUMA-Nachrichten 48: 139-163.
- Ganzeboom, H.B.G./De Graaf, P.M./Treiman, D.J./De Leeuw, J., 1992: A Standard

International Socio-Economic Index of Occupational Status. *Social Science Research* 21: 1-56.

Ganzeboom, H.B.G./Treiman, D.J., 2003: Three Internationally Standardised Measures for Comparative Research on Occupational Status. S. 159-193 in: Hoffmeyer-Zlotnik, J.H.P./Wolf, C. (eds.), *Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables*. New York: Kluwer Academic/Plenum Publishers.

Ganzeboom, H.B.G./Treiman, D.J., 1996: Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations. *Social Science Research* 25: 201-239.

Geis, A.J./Hoffmeyer-Zlotnik, J.H.P., 1991: Zur Vercodung von Beruf, Branche und Prestige für die DDR. S. 139-147 in: *Lebenslagen im Wandel: Basisdaten und -analysen zur Entwicklung in den Neuen Bundesländern*; herausgegeben von: Projektgruppe „Das Sozio-ökonomische Panel“. Frankfurt/Main: Campus.

Hoffmeyer-Zlotnik, J.H.P., 1998: „Beruf“ und „Stellung im Beruf“ als Indikatoren für soziale Schichtung. *RKI-Schriften* 1/1998: 54-64.

Hoffmeyer-Zlotnik, J.H.P./Warner, U., 1998: Die Messung von Einkommen im nationalen und internationalen Vergleich. *ZUMA-Nachrichten* 42: 30-65.

International Labour Office (ed.) 1968: *International Standard Classification of Occupations*. Revised Edition 1968. Geneva: ILO

International Labour Office (ed.) 1990: *ISCO-88. International Standard Classification of Occupations*. Geneva: ILO

Mayer, K.U., 1979: Berufliche Tätigkeit, berufliche Stellung und beruflicher Status - empirische Vergleiche zum Klassifikationsproblem. S. 79-123 in: Pappi, Franz Urban (Hrsg.), *Sozialstrukturanalysen mit Umfragedaten*. Königstein/Ts.: Athenäum.

Treiman, D.J., 1979: Probleme der Begriffsbildung und Operationalisierung in der international vergleichenden Mobilitätsforschung. S. 124-167 in: Pappi, Franz Urban (Hrsg.), *Sozialstrukturanalysen mit Umfragedaten*; Königstein/Ts.: Athenäum.

Treiman, D.J., 1977: *Occupational Prestige in Comparative Perspective*. New York: Academic Press.

Treiman, D.J., 1975: Problems of Concept and Measurement in the Comparative Study of Occupational Mobility. *Social Science Research* 4: 183-230.

Wegener, B., 1988: *Kritik des Prestige*. Opladen: Westdeutscher Verlag.

Wolf, C., 1995: Sozio-ökonomischer Status und berufliches Prestige. *ZUMA-Nachrichten* 37: 102-136.

Anhang

Die zehn Berufe von SIOPS und ISEI jeweils nach ISCO-68 und ISCO-88 mit den höchsten Werten.

Standard International Occupation Prestige Scale (Basis: ISCO-68)

ISCO-68	SIOPS	Berufsangabe
061	78,9	Ärzte
131	78,6	Universitäts- und Hochschullehrer sowie verwandte Lehrkräfte der oberen Bildungsebene
012	76,5	Physiker
122	76,5	Richter
121	73,0	Rechtsvertreter
021	72,8	Architekten und Städteplaner
013	72,3	Physikalisch-wissenschaftliche Berufe, soweit nicht anders klassifiziert
129	71,6	Juristen, soweit nicht anderweitig klassifiziert
063	70,5	Zahnärzte
022	70,3	Bauingenieure

Magnitude Prestige Skala (Basis: ISCO-68)

ISCO-68	MPS	Berufsangabe
061	186,8	Ärzte
122	173,1	Richter
131	167,0	Universitäts- und Hochschullehrer sowie verwandte Lehrkräfte der oberen Bildungsebene
012	163,3	Physiker
129	154,6	Juristen, soweit nicht anderweitig klassifiziert
121	150,8	Rechtsvertreter
219	146,3	Führungskräfte in der Privatwirtschaft, soweit nicht anderweitig klassifiziert
013	140,7	Physikalisch-wissenschaftliche Berufe, soweit nicht anders klassifiziert
063	139,1	Zahnärzte
082	135,4	Mathematiker

International Socio-Economic Index of Occupation Status (Basis: ISCO-68)

ISCO-68	ISEI	Berufsangabe
122	90	Richter
061	88	Ärzte
063	86	Zahnärzte
121	85	Rechtsvertreter
065	84	Tierärzte
002	83	Offiziere
129	82	Juristen, soweit nicht anderweitig klassifiziert
067	81	Apotheker
090	80	Wirtschaftswissenschaftler
012	79	Physiker

Standard International Occupation Prestige Scale (Basis: ISCO-88)

iSCO-88	SIOPS	Berufsangabe
2221	78	Ärzte
2310	78	Universitäts- und Hochschullehrer
2422	76	Richter
2111	75	Physiker und Astronomen
2229	73	Mediziner (ohne Krankenpflege), anderweitig nicht genannt
2421	73	Anwälte
2112	72	Meteorologen
2141	72	Architekten, Stadt- und Verkehrsplaner
1120	71	Leitende Verwaltungsbedienstete
2429	71	Juristen, anderweitig nicht genannt

International Socio-Economic Index of Occupation Status (Basis: ISCO-88)

iSCO-88	ISEI	Berufsangabe
2422	90	Richter
2221	88	Ärzte
1227	87	Produktions- und Operationsleiter in gewerblichen Dienstleistungsunternehmen
2229	85	Mediziner (ohne Krankenpflege), anderweitig nicht genannt
2421	85	Anwälte
2222	85	Zahnärzte
2223	83	Tierärzte
2429	82	Juristen, anderweitig nicht genannt
2213	79	Agrar- u. verwandte Wissenschaftler
2441	78	Wirtschaftswissenschaftler

STUDY MONITORING IN THE INTERNATIONAL SOCIAL SURVEY PROGRAMME (ISSP)

EVI SCHOLZ & SABINE KLEIN

The article discusses study monitoring from a cross-national perspective. It starts from quality standards required for national surveys and then points out what is special for cross-national quantitative surveys. The article stresses the necessary documentation and disclosure of detailed information on cross-national survey methods as a means to evaluate survey quality and enable decisions on comparability. It does not recommend or decide on what is acceptable in terms of comparable methods. The article describes the current state of study monitoring for three important well known cross-national programmes and then goes into more detail of ISSP study monitoring. Study monitoring in the understanding of the authors does not only mean documentation of fielding practice but also promoting trust in survey data.

Der vorliegende Artikel beschäftigt sich mit *study monitoring* aus der Perspektive der international vergleichenden Umfrageforschung. Ausgehend von einer Beschreibung der Qualitätsstandards für nationale Umfragen und der Besonderheiten für international vergleichende Umfragen betont der Beitrag die Bedeutung detaillierter und öffentlich zugänglicher Informationen zur methodischen Vorgehensweise solcher Umfragen. Erst dadurch kann die Qualität einer Umfrage bewertet werden. Dieser Artikel gibt keine Empfehlungen hinsichtlich dessen, was unter vergleichenden Gesichtspunkten methodisch akzeptabel ist, sondern beschreibt den derzeitigen *state of the art*: wie gehen bekannte international vergleichende Umfrageprogramme in Sachen *study monitoring* vor; insbesondere: was bietet ISSP in dieser Hinsicht an. *Study monitoring* geht dem Verständnis der Autoren nach über eine reine Studiendokumentation hinaus und dient durch die Offenlegung von Vorzügen, aber auch Schwächen dazu, Vertrauen in Umfragen zu fördern.

1. Study Monitoring in National and Cross-National Surveys

Study monitoring means documentation of survey methods but also serves to control the quality of a survey. Meaningful for national surveys, study monitoring is more important for cross-national surveys, where quality is dealing with the cross-national implementation and with the comparability of studies across countries. In the following article we discuss study monitoring from a cross-national point of view. We start from quality standards required for national surveys; then we describe the current state of study monitoring for three important well known cross-national programmes to give a brief overview what usually is done with respect to study monitoring. Finally, we go into more detail of the ISSP study monitoring. We do not fix or recommend on what is acceptable in terms of comparable methods. The aim of this article is to stress the necessity of documentation and disclosure of detailed information on cross-national survey methods as a means to evaluate survey quality and to facilitate decisions on comparability. Study monitoring in our understanding is more than a mere documentation of fielding practice but a means of promoting trust in survey data.

1.1 Quality Standards and Study Monitoring

Quality standards, quality control and improvement of quality is an important subject of the scientific discussion of the last years. The Stockholm conference on quality organised by the Swedish Statistical Office in 2001 and the International Conference on Improving Surveys (ICIS) 2002 in Copenhagen are in a line with meetings on non-response in Portland 1999 or ICIS 2000 in Buffalo. A number of publications on quality research for surveys discuss survey quality in the national context in terms of survey measurement and process quality (e.g., Biemer et al. 1991; Lyberg et al. 1997). The American Association for Public Opinion Research (AAPOR) e.g. has published standards of “best practices for survey and public opinion research” asking for quality checks for each stage of the survey¹. AAPOR follows the judgement of the American Statistical Association, and states that the quality of a survey does not depend on size, scope, or prominence of a survey, but on how much attention researchers pay to the problems that can arise at any stage of the survey process including organisation, sampling, questionnaire design, data collection, data processing, and data analysis (ASA 1998).

But defining standards does not automatically result in achieving these standards. Monitoring quality standards in an objective way and reporting the results is therefore important to ensure that these standards are met (Lynn 2001). AAPOR’s demand to dis-

1 http://www.aapor.org/pdfs/best_pra.pdf

close all survey methods and permit evaluation and independent examination is therefore as important as the standards themselves. The catalogue of details to be reported corresponds to the required quality checks (see figure 1). It includes more general information on the purpose of the study or the research, funding, and the fielding institution; it asks for details on the definition of the universe, sample design and sample size²; dates of interviews or fieldwork should be provided as well as full wording of the questions asked; it considers a description of any special editing or data adjustment or indexing procedures as necessary.

Study monitoring thus means documenting and controlling the quality, the development, and the implementation of standards of best practice. Study monitoring is meaningful for national surveys, but it is even more important for cross-national survey research, where quality and quality monitoring is not only dealing with quality in terms of national implementation, but also with the comparability of studies across countries.

1.2 Study Monitoring in Cross-National Surveys

In contrast to the discussion on survey quality for national surveys, in the cross-national context, the discussion on quality mostly deals with questions of equivalence and comparability (e.g. Johnson 1998; van Deth 1998). But though equivalence and comparability are important aspects, they are only necessary but not sufficient conditions for high quality in cross-national surveys. Respecting cultural norms does not excuse lower standards in cross-national studies (Jowell 1998). Methods appropriate for a national survey might not be acceptable in a cross-national one where the additional requirement of comparability produces serious problems, sometimes hard to solve. Differences in sample design, mode of fielding, or non-response might have effects to survey results (Lynn 2001).

The main difference between national survey monitoring and cross-national survey monitoring therefore, from our point of view, results from the additional dimension of comparability. Additional information is desirable taking into account that countries may differ in factors connected to national surveys and their general framework, as the availability and ability of survey research institutions, the availability and coverage of sampling frames, the geographical dispersal of the study population, but also restrictions by country-specific laws or regulations regarding surveys, cultural norms and the language(s) spoken.

Thus, monitoring cross-national surveys becomes more important than for a national survey without comparative aspects. Only a detailed documentation with information on

² Including numbers of non-eligible, not reached, terminations, refusals, and completed interviews.

each individual country's survey methods enable researchers to decide on what they can accept concerning differences between countries (Harkness 1999).

2. Study Monitoring in Practice: The Comparative Study of Electoral Systems, Eurobarometer, World Values Survey

To give an overview of the usual practice and standards on study monitoring of cross-national surveys, we concentrated on the well known survey programmes of Eurobarometer, World Values Survey, and Comparative Study of Electoral Systems. Each programme is concerned with social science research, all are enterprises with large cross-national surveys, and all of them are continuous studies. Of course, this short list is not an all-inclusive one but illustrates the varying level of information and shows how study monitoring is done in present prominent cross-national surveys.

2.1 The Comparative Study of Electoral Systems (CSES)

CSES is a collaborative programme of cross-national election research from all over the world. CSES collects micro-level, public opinion polls data; meso, district level data on electoral returns, turnout, and the number of candidates; and finally, "macro" level data on aggregate electoral returns, electoral rules, and regime attributes. For the micro level data, a common module of questions is included in each CSES member country's election study. These questions, for example, ask about vote choice, evaluation of candidates and parties, of the present and past economic situation, or evaluation of the electoral system.

For quality control and study monitoring, CSES developed and used a questionnaire to describe sampling and data collection. Country specific information – also including the original survey instrument and naming co-operation partners – is published in the CSES pages³ and therefore available for the scientific community's critical review. The methods questionnaire asks about *design*, *process*, and *outcomes*. For each individual country, CSES thus reports on fieldwork mode or as part of *design* description on administration aspects. CSES describes the *process* by details on sample units or on selection methods to identify respondents and informs about substitution, age cut-offs, systematic exclusion of regions or persons, or the interviewers. CSES outlines for example outcomes as fieldwork (length and date), response calculation (refusals, non-contacts, response rate), sample weights or a comparison of the sample to the corresponding national population.

3 For CSES module 1, see <http://www.umich.edu/~cses/download/module1/module1.htm>

2.2 Eurobarometer

Standard Eurobarometer surveys⁴ are the regular representative face-to-face surveys of the European Commission dealing with attitudes towards European integration and the EU of the population of the member countries of the European Union since the 1970s. A report on each individual survey is published, with technical specifications offering some basic information on survey methods annexed⁵. This information includes the names and addresses of co-operating agencies and the responsible research executives, a list of administrative regional units of Europe, and one page on sample specifications. These specifications give short country-specific information on fieldwork (date and length, number of interviews), quasi country-specific information on sampling, and information on weights offered in the data set and used in some published cross-tabulations. Complete question texts for both the basic and the field questionnaires, are available at the pages of the Central Archive at Cologne (ZA)⁶.

2.3 World Values Survey (WVS)

The World Values Survey is a collaborative global programme of socio-cultural and political research. Basic values and beliefs of now almost 80 percent of the world's population are investigated in representative national surveys. In addition to the little information published at the WVS-homepage on methods, sampling, and fieldwork⁷, the codebooks⁸ give some more information.

Details on study monitoring of the three international survey programmes are presented in figure 1. AAPOR's requirements are taken as a guideline to judge the quality of study monitoring, expanded by the additional requirements for cross-national survey monitoring mentioned above. The last column gives the information for ISSP whose approach to monitor its national members' survey methods will be described in the next chapter.

⁴ There are additional surveys by the European Commission on special topics, for example the perception of the EURO; candidate countries Eurobarometer (former Central and Eastern Eurobarometer), and flash Eurobarometer on ad hoc themes.

⁵ Publicly available as: European Commission: Eurobarometer. Public Opinion in the European Union. Report Number ##; also at http://europa.eu.int/comm/public_opinion for the surveys of the last five years.

⁶ http://193.196.10.16/en/data_service/eurobarometer/standard_eb_profiles/indexframe_profiles.htm

⁷ <http://wvs.isr.umich.edu/index.html>

⁸ Available at the Inter-university Consortium for Political and Social Research (ICPSR): <http://www.icpsr.umich.edu>

Figure 1: Study Monitoring in Practice

REQUIREMENTS	CROSS-NATIONAL SURVEY PROGRAMMES			
	CSES	EB	WVS	ISSP
General Information				
Funding, research, fielding institution	•	•	•	•
Purpose of the study	•	•	•	•
Design				
Full wording of the questions asked	•	•	•	•
Definition of the universe	•	•	•	•
Sample design	•	general, not country specific	basic, not country specific	•
Sampling selection procedure	•	general, not country specific	basic, not country specific	•
Outcome of Sample Implementation				
Sample size	•	n of interviews	n of interviews	•
Response rates or number of refusals	•			•
Special data editing				
Interviews				
Dates of interviews or fieldwork	•	•	•	•
Interviewer characteristics	•	included in data set		
Interviewer instructions, codebooks etc.	•	•	•	•
Factors Connected to Nation				
Availability and coverage of sampling frames				
Restrictions by country-specific laws				
Availability and ability of survey research institutions				
Geographical distribution of the study population				
Language(s)	•	included in data set	•	•
Cultural norms				

Source:

CSES on Module 1 with 39 (survey archived for 31) countries: 1996-2000 (required information asked in the original study monitoring questionnaire used by Norway; information from CSES pages);

Eurobarometer 52.0 / 1999 with 15 countries (information from EU and ZA pages);

WVS third wave with 66 countries: 1995-1998 (from WVS pages and ICPSR codebook);

ISSP 1999 with 31 (survey archived for 24) countries: study monitoring published at http://www.gesis.org/Publikationen/Berichte/ZUMA_Methodenberichte/documents/pdfs/tb03_03.pdf

A brief look at figure 1 shows that none of the programmes offers all details of possible interest. Another outcome is that the level and the kind of information on study monitoring for various survey programmes is different: CSES is one of the few cross-national programmes offering open and detailed information on its own survey methods to the public, though the programme does not present a common report to enable an easy view on individual member countries' fielding procedures.

Eurobarometer offers some, rather general information on how the survey is fielded.

Information on methods for WVS is far from being detailed and not persistently country-specific. The level on methodological information offered by and for WVS is rather scant.

3. Study Monitoring in the ISSP

The ISSP is a collaborative programme of social science survey research all over the world. In figure 2, the member countries are listed to give an idea of the different cultural contexts ISSP covers nowadays.

Figure 2: ISSP Member Countries 2002



A module on a topic important for social sciences, such as 'role of government' or 'national identity' is finalised for fielding annually. Modules are implemented and archived in accordance with the ISSP Working Principles and the Central Archive requirements for data sets and background variables. Both of them provide guidelines for individual countries fielding annual modules. These require, for example, national full probability samples and hereby outlaw quota sampling. Members are obliged to comply with these quality standards consistently. Regular control of survey implementation in all ISSP member countries is therefore done by study monitoring in the ISSP. Conducting special surveys for quality monitoring of the ISSP was first agreed on at the 1996 general assembly of the ISSP. In 1999, the general ISSP assembly decided on satisfactory completion of the study monitoring questionnaire as obligatory condition for including a national data set into the merged ISSP data file. Starting with the 1995 monitoring study by Park and Jowell (1997) and continued by ZUMA from then on, the monitoring reports are supplemented to the ISSP codebooks and are published at the German Central Archive's web site⁹, which also is the ISSP archive. The ISSP hereby offers detailed information on its methods not only for researchers inside the ISSP itself but also for any interested researcher and user. Study monitoring surveys collect information on *design*, *process*, and *outcome* of individual implementations of ISSP modules. The resulting reports are quite short and neutral, do not judge but mainly chart important aspects at the national survey level. For the time being, the study monitoring report is available on ISSP 1995 (National Identity), ISSP 1996 (Role of Government), ISSP 1997 (Work Orientations), ISSP 1998 (Religion), and ISSP 1999 (Social Inequality). Study monitoring is currently in work for ISSP 2000 (Environment), ISSP 2001 (Social Networks), and ISSP 2002 (Family and Changing Gender Roles).

Study monitoring in the ISSP serves several aims: The first one is to establish quality monitoring on a systematic basis. A second one is to monitor and evaluate the extent to which individual members adhere to ISSP implementation requirements. A third aim is to expand and improve the documentation available for all working with ISSP data. Based on the internal ISSP monitoring information, users of ISSP data are able to decide on the quality and the comparability of given components across countries. For cross-national surveys, such a detailed information is missing too often. Collecting and publishing this information, the ISSP becomes one of the few cross-national survey programmes with transparency on methods.

The monitoring questionnaire developed originally by the National Centre for Social Research (formerly SCPR) asks more detailed questions than the study description sheet

⁹ http://www.gesis.org/en/data_service/issp/data/list_cdbk_pdf.htm

already used from the beginning of the ISSP. In the following years the questionnaire was modified continuously. One of these modifications was the development of two specially tailored questionnaires, one for face-to-face (and self-completion) surveys and one for mail surveys.

Park and Jowell (1997) on ISSP 1995 and then Harkness et al. (2001, 2003) on the following ISSP surveys starting with 1996 called attention to several aspects and differences in the ISSP studies in the three main areas of *design*, *process*, and *outcomes*. Some, of course, are more important and critical than others. Some problems on comparability can or will be solved, others will probably remain: Quota sampling being not best practice may be substituted by a full probability sample if severe efforts are undertaken. The problem of non-response increases with variation between countries (Svallfors 1999) and therefore, ISSP study monitoring surveys gather as much information as possible about the respective non-respondents or refusals to tackle the problem. Other country-specific procedures may result from legal restrictions or financial constraints. Full homogeneity in the enterprise of cross-national survey research is hardly available.

ISSP study monitoring questionnaires ask for the following:

3.1 Design

By *design*, we refer to all aspects of implementation of the annual ISSP module itself, such as translation, mode(s) and context of administration, and questions.

- *Translation* – One of the major concerns in international comparative surveys is the aspect of translation. Apart from the general requirement of translating the British source questionnaire, several countries have to deal with multiple translation for the cultural and linguistic composition of their countries. Several countries have to translate into two or even more languages, such as Israel into Hebrew and Arab, or Switzerland where a German, French, and an Italian field questionnaires are used. Information is needed on who actually did the translation(s). It can, for example, be done in-house (within the institution responsible for the national ISSP), by outside experts or by a combination of both. Other questions deal with the topics of evaluation and pre-testing of the translated questionnaires. Both of them are voluntary, but in terms of quality highly recommendable.
- *Mode of administration* – there are several options how to field the module. One can opt for face-to-face, self-completion with some interviewer involvement, or self-completion by mail. Closely connected to this is the context in which the module is

fielded, either as a survey on its own or part of a larger survey. (With the latter, self-completion with some interviewer involvement is frequently met.)

Depending on the respective mode, the questionnaire asks about the interviewer performance, management, and control or mail survey procedures (actual number of mailings and material that is sent out in each mailing).

Most of the countries of ISSP conduct face-to-face interviews. In 1999, it was more than the half of the 24 archiving countries¹⁰. About the same proportion uses either self-completion with some interviewer involvement or self-completion by mail. The distribution shown in figure 3 indicates a slight bias on self-completion for the Western world whereas most of the Eastern European countries decided to proceed a face-to-face survey.

Figure 3: Different Modes in the ISSP-Module on 'Social Inequality' (1999)

<i>Self-Completion</i>	<i>Self-Completion by Mail</i>	<i>Face-to-Face</i>	
Germany	Australia	Austria	Latvia
Great Britain*	Canada	Bulgaria	Portugal
Japan	France	Chile	The Philippines
Poland	Norway	Cyprus	Russia
USA	New Zealand	Czech Republic	Slovenia
	Sweden	Hungary	Spain
		Israel	

* including Northern Ireland

- *Questions* – ISSP modules have a prescribed question order for all substantive questions, most of these (called core items) are obligatory. The study monitoring questionnaires ask whether member countries' questionnaires comply with the prescribed order and whether all the core items are included.

3.2 Process

Since the sampling standards differ from country to country, a major part of the questionnaire is dedicated to sampling. Information is asked on the respective target population (e.g., whether foreigners are part of it), the sampling procedures (e.g., multistage; stratified), the sampled population (e.g., whether anyone is excluded), and finally the sampled unit (individual, address, household). Depending on the sampled unit, another selection

¹⁰ From the 31 ISSP member countries in 1999, 24 countries having archived their data successfully.

criterion has to be identified, such as Kish grid or the birthday method. Other questions deal with the use of quota and substitution procedures.

Which groups of persons are part of the population sampled varies between countries. For the 1999 module of ISSP, e.g., Austria, Chile, Poland, Russia, and Slovenia included only citizens of their countries while all others also included adults of any other nationality. If the proportion of minorities is very low, this aspect might be of minor relevance. However, for Austria EUROSTAT reports about 9 percent foreigners of the total population in 1998, with an overwhelming majority not originating from EU countries (EUROSTAT 2001, 30f.).

Another aspect with regard to the target population is whether to include persons living in institutional accommodation, such as old age homes or prisons. In the 2000 ISSP, there are only four countries including persons living in institutional accommodations into their population sampled, namely Denmark, Norway, New Zealand, and Sweden. For Germany, surveying only the private household population, the estimated proportion of elderly in institutional accommodation is about 3 percent to 4 percent (Schneider 1998).

In this context, another question of importance arises: did countries adopt an upper age cut-off or not? Undoubtedly population is growing older and older; and consequently the proportion of the elderly in a country's total population is ever increasing. Countries using an upper age cut-off in the ISSP 2000 are Denmark (upper age cut-off 74), Norway (upper age cut-off 79) and Sweden (upper age cut-off 79). The share of the adult population older than the age cut-off is 5.5 percent for Norway, 6.2 percent for Sweden and 8.8 percent for Denmark (source IDB)¹¹. Figure 4 shows the percentages of respondents in selected countries following the age cut-offs used by Denmark and Norway or Sweden. As could be seen, the share of elderly in the national population differs cross-nationally. For some but not many, the share is quite low but for most of the ISSP countries, old-age persons are not a negligible part of countries' total populations.

Though ISSP working rules do not include restrictions on an upper age cut-off, it is already one of the items of the working list of the ISSP methodology committee.

¹¹ The International Data Base (IDB) is a publicly accessible and computerized source of demographic and socio-economic statistics for 227 countries and areas of the world; IDB estimations and statistics are based on the data of the U.S. Bureau of Census, National Statistical Offices, and UN agencies, especially censuses and surveys.

Figure 4: ISSP Respondents (ISSP 2000) of Selected Countries without Upper Age Cut-Off in Respective Age Categories and Proportion of Total Population (IDB)

	<i>Age 75 plus (%)</i>		<i>Age 80 plus (%)</i>	
	ISSP	IDB	ISSP	IDB
Austria	11.6	8.5	6.5	4.1
Canada	11.4	7.2	8.3	3.9
Spain	11.9	8.4	6.1	4.3
Czech Republic	2.7	6.6	0.4	2.8
Latvia	2.6	6.9	0.8	3.5
The Philippines	1.7	1.8	0.7	0.8

Source: data on ISSP 2000 based on own calculation; for proportion of total population, see International Data base

3.3 Outcomes

- *Response and Outcome Figures* – Ever increasing attention focuses on response or outcome figures. Thus, detailed documentation of response and non-response is necessary to allow for assessment of survey quality. Here we ask for all available figures relating to the issued sample, refusals, non-respondents, and completed interviews or returned questionnaires; for example the number of selected respondents too sick or incapacitated to participate, the number of selected respondents with inadequate understanding of language of survey, the number of refusals at the selected address, the number of proxy refusals (meaning the number of refusals on behalf of the selected respondent) or the number of personal refusals by the selected respondent him/herself.
- *Date and length of fieldwork* – an annual ISSP module is to be fielded within a specified and limited period of time. Since the length of fieldwork does vary, we ask for it.
- *Data checking* – This part of the questionnaire deals with establishing the extent to which the respective data was checked. The ISSP study monitoring questionnaires ask, for example, whether data are checked for consistency, whether data are checked to ensure they fell within permitted ranges, whether filter instructions were followed correctly, whether errors were corrected individually or automatically.

4. Conclusion

When designing a survey or when analysing survey data, various drawbacks or traps need to be taken into account, such as effects resulting from (the quality of) sampling procedures, question wording, or the appropriateness of the mode used. True for a national survey, it is all the more true for a comparative survey in a cross-national context. Here the need for a proper survey documentation and detailed information is even more conspicuous, as there might be effects, e.g., from the translation of questions or varying sampling procedures between countries. This is what study monitoring in cross-national surveys is about, collecting and documenting information about any relevant aspect of each national part of the survey. Study monitoring hereby is not only a means to support and inform data users but it may also serve as a starting point for quality control as well. And, finally, provided with such information, researchers have the chance to decide on comparability. Study monitoring in the end means promoting trust in the use of survey data.

Korrespondenzadresse

Evi Scholz
ZUMA
Postfach 12 21 55
68 072 Mannheim
email: scholz@zuma-mannheim.de

References

- American Association for Public Opinion Research (AAPOR) 2002: Best Practices for Survey and Public Opinion Research (available online www.aapor.org/ethics/best.html).
- American Statistical Association 1998: Judging the Quality of a Survey. ASA: Section on Survey Research Methods.
- Biemer, P. P./Groves, R.M./Lyberg, L.E./Mathiowetz, N.A./Sudman, S. (eds.) 1991: Measurement Errors in Surveys. New York: John Wiley and Sons.
- EUROSTAT 2001: Beschreibung der sozialen Lage in Europa, (available online http://europa.eu.int/comm/eurostat/Public/datashop/print-product/DE?catalogue=Eurostat&product=KE-36-01-702-__-N-DE&mode=download).
- Harkness, J., 1999: In Pursuit of Quality: Issues for Cross-national Survey Research, *International Journal of Social Research Methodology*, 2: 125-140.
- Harkness, J./Klein, S./Scholz, E. 2003: ISSP Study Monitoring 1999: Reports to the ISSP General Assembly on monitoring work undertaken for the ISSP by ZUMA, ZUMA-Metho-

denbericht 2003/3. Mannheim: ZUMA (available online www.gesis.org/Publikationen/Berichte/ZUMA_Methodenberichte/documents/pdfs/tb03_03.pdf).

Harkness, J./Langfeldt, B./Scholz, E. 2001: ISSP Study Monitoring 1998, Reports to the ISSP General Assembly on monitoring work undertaken for the ISSP by ZUMA, Germany (available online www.za.uni-koeln.de/data/en/issp/codebooks/s3190app.pdf).

Harkness, J./Langfeldt, B./Scholz, E. 2001: ISSP Study Monitoring 1997, Reports to the ISSP General Assembly on monitoring work undertaken for the ISSP by ZUMA, Germany (available online www.za.uni-koeln.de/data/en/issp/codebooks/s3090app.pdf).

Harkness, J./Langfeldt, B./Scholz, E. 2001: ISSP Study Monitoring 1996, Reports to the ISSP General Assembly on monitoring work undertaken for the ISSP by ZUMA, Germany (available online www.za.uni-koeln.de/data/en/issp/codebooks/s2900app.pdf).

Johnson, T., 1998: Approaches to Equivalence in Cross-Cultural and Cross-National Survey Research, in: Harkness, J. (ed.), Cross-Cultural Survey Equivalence, ZUMA-Nachrichten Spezial, 3: 1-40.

Jowell, R., 1998: How Comparative is Comparative Research? Working Paper of the Centre for Research into Elections and Social Trends, 88, September 1998.

Lyberg, L./Biemer, P./Collins, M./de Leeuw, E./Dippo, C./Schwarz, N./Trewin, D. (eds.) 1997: Survey Measurement and Process Quality. New York: John Wiley and Sons.

Lynn, P., 2001: Developing Quality Standards for Cross-National Survey Research: Five Approaches, Working Papers of the Institute for Social and Economic Research, paper 2001-21. Colchester: University of Essex.

Park, A./Jowell, R. 1997: Consistencies and Differences in a Cross-national Survey. The International Social Survey Programme (1995) (available online www.za.uni-koeln.de/data/en/issp/codebooks/s2880app.pdf).

Schneider, S., 1998: Ältere Bundesbürger in Privathaushalten und in Heimen - Lebenssituation und Heimeintrittsgründe - Repräsentative Ergebnisse für die Bundesrepublik Deutschland, Sozialer Fortschritt 47: 30-37.

Svallfors, S., 1999: National Differences in National Identities. An introduction to the International Social Survey Programme in: Toš, N./Mohler, P.Ph./Malnar, B. (eds.), Modern Society and values, Ljubljana: Faculty of Social Sciences, (first published in New Community, 22: 127-34): 3-13.

U.S. Census Bureau 2000: International Data Base, IDB Summary Demographic Data, (available online <http://www.census.gov/ipc/www/idbsum.html>).

Van Deth, J.W., 1998: Equivalence in Comparative Political Research, in: van Deth, J. W. (ed.), Comparative Politics. The Problem of Equivalence, London and New York: Routledge: 1-19.

MIKRODATEN (GERMAN MICRODATA LAB): DAS SERVICEZENTRUM FÜR AMTLICHE MIKRODATEN BEI ZUMA

*PAUL LÜTTINGER, BERNHARD SCHIMPL-NEIMANNS
HEIKE WIRTH & GEORG PAPASTEFANOU*

Am 1. Juli 2003 werden die beiden ZUMA-Abteilungen Mikrodaten und Einkommen & Verbrauch in eine neue organisatorische Einheit überführt. Damit nimmt das vom Gründungsausschuss des Rates für Sozial- und Wirtschaftsdaten vorgeschlagene „Servicezentrum für Mikrodaten der GESIS bei ZUMA“ seine Arbeit auf. Gemeinsam mit dem Internationalen Datenservicezentrum am Forschungsinstitut zur Zukunft der Arbeit (IZA, Bonn) und den Forschungsdatenzentren des Statistischen Bundesamtes (Wiesbaden, Bonn), der Statistischen Landesämter und der Bundesanstalt für Arbeit (Nürnberg) entsteht eine Forschungs- und Servicestruktur für amtliche Mikrodaten, die vor wenigen Jahren noch undenkbar erschien. Ziel der Einrichtungen ist es, für die empirisch arbeitenden Wirtschafts- und Sozialwissenschaften Mikrodaten der amtlichen Statistik zu erschließen, bereitzustellen und einen umfassenden Service zu den Daten anzubieten.

Damit soll die positive Entwicklung in den letzten Jahren, in denen der Zugang zu Mikrodaten der amtlichen Statistik in Form von Scientific Use Files dank der Initiativen des Bundesministeriums für Bildung und Forschung (BMBF), des Statistischen Bundesamtes und der Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen (GESIS) institutionalisiert und wesentlich erleichtert¹ wurde, konsolidiert werden.

¹ Dementsprechend stieg die Zahl der Nutzer solcher Daten in den letzten Jahren sprunghaft an. So arbeiten z.B. bereits laut Auskunft des Statistischen Bundesamtes (Stand vom Februar 2003) 170 Institute zwischenzeitlich in den verschiedensten Forschungskontexten mit Scientific Use Files des Mikrozensus, wobei die Zahl der Wissenschaftler, die mit den Daten arbeiten, erfahrungsgemäß um einiges höher liegen dürfte. Über 100 wissenschaftliche Beiträge - soweit uns bekannt - wurden auf der Basis dieser Daten bereits veröffentlicht und viele weitere Beiträge sind in Arbeit. Die Publikationen sind doku-

Die Einrichtung von Forschungsdaten- und Servicezentren geht auf Empfehlungen der Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) zurück, die vom BMBF eingesetzt wurde. Neben Eigenleistungen der statistischen Ämter und Institute erfolgt die Finanzierung der Zentren durch das BMBF.² Im Folgenden geben wir einen Überblick über die Entwicklung in den letzten Jahren und die neu entstehenden Einrichtungen und stellen die Aufgaben und Arbeiten bei ZUMA vor.

1. Wie es dazu kam

Walter Müller und Richard Hauser haben bereits 1987 ausführlich begründet, warum die routinemäßig veröffentlichten Daten oder auch Sondertabellen der statistischen Ämter wissenschaftlichen Zwecken nicht genügen und Individualangaben (Mikrodaten) für die Forschung benötigt werden (Müller/Hauser 1987). Dies war noch bevor Daten des Mikrozensus 1989 erstmals als Scientific Use File verfügbar waren. Zuvor konnten nur einzelne Forscher - sehr erfolgreich und mit einem hohen wissenschaftlichen Ertrag - amtliche Individualdaten nutzen. Im folgenden Überblick werden die wichtigsten Phasen des Zugangs und der Nutzung amtlicher Mikrodaten nachgezeichnet (vgl. Müller 1999).

Die Verwendung von amtlichen Mikrodaten für die Bearbeitung von Forschungsfragen begann, als ab den siebziger Jahren in Folge der Entwicklung der Datenverarbeitung zunehmend auch in der akademischen Forschung die technischen Voraussetzungen gegeben waren, die umfangreichen Mikrodaten der amtlichen Statistik zu verarbeiten. Vorreiter in der Nutzung solcher Daten waren die Projekte SPES und VASMA sowie der Sonderforschungsbereich 3, in denen eine Vielzahl von Forschungsarbeiten entstanden.³ Der bis zu diesem Zeitpunkt relativ einfache Zugang zu amtlichen Mikrodaten wurde im Zusammenhang mit öffentlichen Diskussionen über Datenschutz ab Ende der siebziger Jahre deutlich erschwert. Nach dem Bundesstatistikgesetz von 1980 wurden Individualda-

mentiert und unter http://www.gesis.org/Dauerbeobachtung/Mikrodaten/Publikationen/MZ_Literatur.htm abrufbar.

² Die Projektfinanzierung beim Servicezentrum für Mikrodaten beginnt voraussichtlich Mitte 2003 und ist zunächst auf drei Jahre befristet. Zum Zeitpunkt des Redaktionsschlusses ist ebenfalls noch offen die Projektfinanzierung des Internationalen Datenservicezentrums am Forschungsinstitut zur Zukunft der Arbeit (IZA), des Forschungsdatenzentrums der Statistischen Landesämter und des Forschungsdatenzentrums der Bundesanstalt für Arbeit.

³ Die ersten großen Forschungsprojekte, die mit amtlichen Mikrodaten arbeiteten, waren das DFG-Projekt „Sozialpolitisches Indikatoren- und Entscheidungssystem“ (SPES; 1972-1978) und der Sonderforschungsbereich 3 (1979-1990), die gemeinsam von den Universitäten Frankfurt und Mannheim durchgeführt wurden. Das von der Stiftung Volkswagenwerk geförderte Projekt „Vergleichende Analyse der Sozialstruktur mit Massendaten“ (VASMA) wurde 1979-1984 an der Universität Mannheim durchgeführt (siehe <http://www.sowi.uni-mannheim.de/lehrstuehle/lesas/forschung/vasma/vasma.html>).

ten von den statistischen Ämtern nur noch in Form stark aggregierter Tabellen weitergegeben, deren Analysemöglichkeiten erheblich eingeschränkt waren.⁴

Erst mit der Novellierung des Bundesstatistikgesetzes (BStatG) 1987 wurden die rechtlichen Grundlagen für eine verbesserte Datenweitergabe geschaffen. Nach § 16 Abs. 6 BStatG können für die Durchführung wissenschaftlicher Vorhaben von den statistischen Ämtern Einzelangaben an Forschungseinrichtungen übermittelt werden, wenn sie Auskunftgebenden oder Betroffenen nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden können. Zur Operationalisierung dieses Konzepts der faktischen Anonymität wurden im Anonymisierungsprojekt (Müller et al. 1991, Kooperation zwischen Universität Mannheim, Statistischem Bundesamt, Statistischen Landesämtern und ZUMA) konkrete Empfehlungen entwickelt, welche die Voraussetzungen für die Bereitstellung von Scientific Use Files des Mikrozensus (ab 1989) und anderer Daten waren. Dennoch wurden die Mikrozensusdaten nur von wenigen Forschungseinrichtungen genutzt, da von der amtlichen Statistik sehr hohe Kosten für die Bereitstellung der Daten erhoben werden mussten. Dies führte zu einer Initiative von Wolfgang Zapf, dem damaligen Vorsitzenden des Kuratoriums der Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen (GESIS), und Walter Müller, der sich weitere Forschungseinrichtungen und einzelne Forscher anschloss.⁵ Nach Gesprächen mit dem Statistischen Bundesamt und dem Bundesministerium für Bildung und Forschung (BMBF) konnten ab Ende 1996 insbesondere durch die finanzielle Unterstützung des BMBF die Kostenprobleme für die Mikrozensus 1995 und 1996 gelöst und weitere Scientific Use Files zugänglich gemacht werden.⁶ Darüber hinaus haben die GESIS und das Statistische Bundesamt für vier weitere Mikrozensusdaten einen kostengünstigen Datenzugang vereinbart, wobei die GESIS die Grundfinanzierung der Datenbereitstellungskosten übernommen hat. All dies hat bewirkt, dass das Potenzial amtlicher Mikrodaten durch die Sozial- und Wirtschaftsforschung seither breiter als zuvor genutzt

⁴ In §11, Abs. 5 BStatG 1980 wurde festgehalten, dass Einzelangaben dann weiter gegeben werden dürfen, wenn sie Auskunftspflichtigen oder Betroffenen nicht mehr zuzuordnen sind. Im Absatz 6 war geregelt, dass eine Zusammenfassung von Angaben mehrerer Auskunftspflichtiger nicht als Einzelangabe zählt. Obwohl beide Absätze damals mit dem Ziel aufgenommen wurden, eine Weitergabe von Daten an die Forschung möglich zu machen, führte die rechtliche Auslegung von Seiten der amtlichen Statistik sowie die Einschätzung, unter welchen Bedingungen eine Reidentifikation zweifelsfrei ausgeschlossen werden kann (m.a.W. wann eine „absolute“ Anonymisierung vorliegt), in der Praxis zu sehr restriktiven Regelungen der Datenweitergabe.

⁵ Siehe hierzu ZUMA-Nachrichten Nr. 39, 1996: 171-178.

⁶ So die Einkommens- und Verbrauchsstichprobe 1993, Deutsche Stichprobe des Europäischen Haushaltspanels 1994-1996, Zeitbudgeterhebung 1991/92; siehe ZUMA-Nachrichten Nr. 41 1997: 182-187; vgl. auch Köhler et al. 2000.

wird. Eine Übersicht über die gegenwärtig verfügbaren Scientific Use Files findet sich im Anhang (Tabelle 1).⁷

Um den Zugang zu kostengünstigen Mikrodaten zu institutionalisieren und darüber hinaus grundlegende Fragen des Verhältnisses zwischen amtlicher Statistik und Wissenschaft zu klären, die im Memorandum zur Situation der empirischen Wirtschaftsforschung (Hauser et al. 1998) thematisiert wurden, wurde 1999 ein Symposium veranstaltet (Statistisches Bundesamt 1999), das in „Thesen zur Verbesserung der informationellen Infrastruktur“ resultierte. Zudem wurde die Einrichtung einer Kommission empfohlen, die danach von der Bundesministerin für Bildung und Forschung mit dem Auftrag bestellt wurde, Lösungsvorschläge zu einem intensiveren Zusammenwirken von Wissenschaft und Statistik in Bezug auf inhaltliche, institutionelle und ressourcenbezogene Fragestellungen zu erarbeiten. Diese „Kommission zur Verbesserung der informationellen Infrastruktur“ (KVI), in der Vertreter der amtlichen Statistik bzw. der öffentlichen Datenproduzenten und Vertreter der Sozial- und Wirtschaftswissenschaften versammelt waren, hat eine Reihe von Vorschlägen erarbeitet, die von institutionellen Regelungen der Kooperation über Empfehlungen zu einzelnen Datenerhebungen, Methodenforschung, Archivierung und Datenschutz reichen,⁸ wobei im Mittelpunkt Aspekte des Zugangs zu Mikrodaten stehen (KVI 2001).⁹ Als einen ersten Schritt zur Institutionalisierung der Zusammenarbeit zwischen Wissenschaft und Statistik hat das BMBF einen Ausschuss zur Gründung eines „Rates für Sozial- und Wirtschaftsdaten“ berufen.¹⁰ Dieser Gründungsausschuss hat auf der Grundlage der KVI-Anträge zur

⁷ Zu den Zugangsmöglichkeiten zu Scientific Use Files und auch zu Public Use Files sowie zu Daten, die über Ferndatenanalysen zugänglich sind, siehe http://www.gesis.org/Dauerbeobachtung/Mikrodaten/Informationsquellen/weitere_Mikrodaten.htm.

⁸ Die Empfehlungen der KVI zur Verbesserung der Weiterbildung insbesondere im Bereich von Methoden aufgreifend, plant das Zentralarchiv eine „Interdisziplinäre Summer School für empirische Sozial- und Wirtschaftsforschung“ und hat für die Durchführung einer Machbarkeitsstudie Projekt-mittel beantragt. Zu diesem Zusammenhang der KVI-Empfehlungen und der Methodenausbildung siehe auch Müller 2001.

⁹ Die Veröffentlichung enthält auf einer beigelegten CD-ROM die für die KVI erstellten Expertisen. Für eine Zusammenfassung der Kommissionsarbeiten siehe Pressemitteilung des BMBF vom 13. März 2001 unter: <http://www.bmbf.de/presse01/338.html>. Eine Kurzfassung des KVI-Gutachtens ist unter http://www.bmbf.de/presse01/A--FIN4_.pdf zugänglich.

¹⁰ Die Mitglieder des Gründungsausschusses sind: Präsident Johann Hahlen (Statistisches Bundesamt), Prof. Dr. Richard Hauser (Universität Frankfurt), Präsident Eckart Hohmann (Hessisches Statistisches Landesamt), Prof. Dr. Wolfgang Jagodzinski (Universität Köln), Prof. Dr. Gerhard Kleinhenz (Universität Passau), Prof. Dr. Hans-Jürgen Krupp (Universität Frankfurt, stellvertretender Vorsitzender), Prof. Dr. Karl Ulrich Mayer (MPI für Bildungsforschung, Berlin), Prof. Dr. Walter Müller (Universität Mannheim), Ltd. Verw. Direktor Uwe Rehfeld (Verband Deutscher Rentenversi-

Einrichtung von Forschungsdatenzentren bei den Datenproduzenten sowie von in der Forschung verankerten Servicezentren angeregt, evaluiert und dem BMBF zur Förderung empfohlen.

Zusammenfassend kann man zu dieser erfreulichen Entwicklung in den letzten Jahren festhalten, dass viele Verbesserungen im Zugang und der Nutzung amtlicher Mikrodaten in Gang gesetzt sind und in der nächsten Zeit wirksam werden. Auf die konkreten Arbeiten bei ZUMA wollen wir im Folgenden näher eingehen.

2. Mikrodaten (GML) bei ZUMA

Die Aktivitäten von ZUMA zu amtlichen Mikrodaten gehen auf eine Initiative aus der Wissenschaft und einer darauf folgenden Empfehlung des Wissenschaftsrates von 1986 zurück, durch die GESIS beauftragt wurde, bei ZUMA eine Mikrodatenabteilung als Nachfolgeeinrichtung des VASMA-Projektes einzurichten und forschungsgerechte Nutzungsmöglichkeiten für amtliche Mikrodaten zu schaffen. Ziel war es, diese damals für die empirische Wirtschafts- und Sozialforschung teilweise nur schwierig, teilweise überhaupt nicht nutzbaren Daten zu erschließen, als festen Bestandteil der deutschen Forschungsinfrastruktur zu verankern und für die effiziente Nutzung in der Forschung den notwendigen wissenschaftlichen Service auf der Basis eigener wissenschaftlicher Arbeiten aufzubauen. Diese Aufgabenstellung von in der Wissenschaft verankerten Serviceeinrichtungen hat die KVI in ähnlicher Weise begründet: „Es ist wichtig, dass die Organisation und Erledigung der Servicefunktionen wissenschaftsnah erfolgt, damit die Entwicklungen in der Wissenschaft rechtzeitig erkannt und die daraus sich ergebenden Daten- und Servicebedarfe daran ausgerichtet werden. Gleichzeitig ist eine Unabhängigkeit von den Datenproduzenten notwendig, um die Entwicklungen dort mit einer gewissen kritischen Distanz zu verfolgen, die Interessen der Wissenschaft zu bündeln und gegenüber den Datenproduzenten zu vertreten“ (KVI 2001: 34).

2.1 Service – wissenschaftliche Dienstleistungen

ZUMA konzentriert sich bei seinen wissenschaftlichen Dienstleistungen auf die beiden großen Haushaltsstichproben der amtlichen Statistik, den Mikrozensus und die Einkommens- und Verbrauchsstichprobe (EVS).¹¹ Beide Datenbestände zeichnen sich durch ein

cherungsträger), Prof. Dr. Gert G. Wagner (DIW Berlin, Vorsitzender), Prof. Dr. Joachim Wagner (Universität Lüneburg).

¹¹ Diese Daten werden ergänzt durch ausgewählte amtliche Mikrodaten der DDR-Statistik und die vom VASMA-Projekt übernommenen historischen Mikrodaten. Ausserdem ist ZUMAs wissen-

hohes Potenzial zur Analyse von für die Sozial- und Wirtschaftswissenschaften zentralen Fragestellungen aus, unter anderen zum Arbeitsmarkt, zur demografischen und sozialen Struktur der Gesellschaft, zur Einkommensverteilung und -verwendung oder zu den Formen des Zusammenlebens. Sie bieten aufgrund ihrer regelmäßigen Periodizität mit weitgehend identischen Frageprogrammen interessante Möglichkeiten längsschnittlich angelegter Auswertungen zu Fragen des sozialen und wirtschaftlichen Wandels auf der Basis langer Datenreihen. Neben dem Schwerpunkt des wissenschaftlichen Service zu Scientific Use Files des Mikrozensus und der EVS werden übergreifende Funktionen in Bezug auf amtliche Mikrodaten und deren Erschließung für die gesamte Sozial- und Wirtschaftsforschung wahrgenommen.

Im Zusammenhang mit den Empfehlungen der KVI zum *Ausbau der Dateninfrastruktur* setzt ZUMA in Zusammenarbeit mit der amtlichen Statistik kurz- und mittelfristig zwei Schwerpunkte für die Bereitstellung neuer Scientific Use Files:

- *Erweiterung der Nutzungsmöglichkeiten des Mikrozensus sowie der Einkommens- und Verbrauchsstichprobe in inhaltlicher und zeitlicher Perspektive.*
Zur Deckung des Bedarfs der Forschung an weiteren Daten wird die Bereitstellung von älteren Erhebungen der Einkommens- und Verbrauchsstichprobe (vor 1993) und des Mikrozensus (vor 1989) angestrebt. Des Weiteren sollen beim Mikrozensus die Möglichkeiten der Bereitstellung von Regionalfiles und Panelfiles ausgeschöpft werden.
- *Unterstützung von Projekten zur Öffnung des Zugangs zu Daten, für die bisher noch keine Anonymisierungskonzepte vorliegen.*
Zur Erschließung neuer Datenbasen wird ZUMA gemeinsam mit den Datenbesitzern und Forscher/innen in Projekten an Konzepten für Scientific Use Files mitarbeiten und hierbei seine Expertise insbesondere in Form von Beratungen zu verschiedenen Fragen der Umsetzung faktischer Anonymität einbringen. So unterstützt ZUMA zurzeit das vom BMBF geförderte Projekt „Faktische Anonymisierung wirtschaftsstatistischer Einzeldaten“ der statistischen Ämter und des Instituts für Angewandte Wirtschaftsforschung (IAW),¹² (siehe Wirth 2003). Weitere Projekte werden folgen.

Neben diesen auf die Primärererschließung der Daten gerichteten generellen Dienstleistungen bietet ZUMA einen umfassenden spezifischen *Service für die Nutzer der Scientific Use Files* des Mikrozensus und der EVS an. Seit dem Abschluss des Pilotprojekts, mit

schaftlicher Service auf die Erschließung von Verbraucherpaneldaten gerichtet, wie sie z.B. in den Haushalts- und Individualpanels der GfK vorliegen (siehe Papastefanou 2002).

¹² Am 20./21. März 2003 fand der 1. Workshop zur Anonymisierung wirtschaftsstatistischer Einzeldaten statt. Sie hierzu <http://www.uni-tuebingen.de/iaw/iaw.fawe-nutzertagung.html>.

dem auch neue Formen der Zusammenarbeit zwischen Statistik und Wissenschaft erprobt werden sollten, wird dieser Service in enger Zusammenarbeit mit dem Statistischen Bundesamt erbracht. Bei diesen spezifischen Angeboten geht es vor allem darum, den Aufwand des Einzelforschers bei der Informationsbeschaffung, der Erarbeitung des datenbezogenen methodischen und inhaltlichen Grundlagenwissens, der Datenaufbereitung und der Datenanalyse zu minimieren und damit die Effizienz der Datennutzung zu steigern. Zu nennen sind u.a.:

- die Entwicklung von Konzepten zur Filebereitstellung und Gespräche mit den statistischen Ämtern zu deren Umsetzung;
- Konsistenzprüfungen der Scientific Use Files; diese ergänzen die Datenprüfungen durch die statistischen Ämter, so dass schon im Vorfeld der eigentlichen Datenbereitstellung durch das Amt Datenprobleme erkannt und behoben werden können;
- die Datenaufbereitung und Datendokumentation entsprechend sozialwissenschaftlicher Standards, wie z.B. Bereitstellung von Setups zur Erstellung von Files im dokumentierten Format, Randverteilungen zur Prüfung der eingelesenen Rohdaten sowie von Metadaten im WWW (Fragebogen, Schlüsselverzeichnis, Klassifikationen und Definitionen der amtlichen Statistik);¹³
- die Umsetzung forschungsüblicher Konzeptualisierungen, Klassifikationen und Skalen und damit einhergehend die Bereitstellung der hierfür notwendigen Codierungsroutinen für theoriebasierte sozialwissenschaftliche Klassifikationen oder Skalen wie auch sonstiger Hilfsmittel für die Datenanalyse (Mikrodaten-Tools);¹⁴
- die individuelle Nutzerberatung zu Datennachweisen und Analysemöglichkeiten aber auch spezifischen Problemen bei der Verwendung von amtlichen Mikrodaten;
- Sonderauswertungen für externe Forscher sowie die Ermöglichung von Arbeitsaufenthalten als Gastwissenschaftler bei ZUMA für eigene Datenanalysen;
- Weiterbildungsangebote (Workshops, Nutzerkonferenzen), wobei insbesondere die Nutzerkonferenzen zugleich den Erfahrungsaustausch zwischen den Nutzern und den Austausch zwischen Datennutzern und Datenproduzenten fördern.¹⁵ In Kooperation mit dem Statistischen Bundesamt wird im Oktober 2003 die dritte Mikrozensus-Nutzerkonferenz stattfinden (siehe das Tagungsprogramm S. 179ff. in diesem Heft).

¹³ Siehe zum Mikrozensus <http://www.gesis.org/Dauerbeobachtung/Mikrodaten/Daten/Abteilungsdaten/Mikrozensus/mz.htm>.

¹⁴ Siehe http://www.gesis.org/Dauerbeobachtung/Mikrodaten/mikrodaten_tools/index.htm.

¹⁵ Die Beiträge der ersten beiden Mikrozensus Nutzerkonferenzen sind dokumentiert in Lüttinger (1999) sowie unter <http://www.gesis.org/Dauerbeobachtung/Mikrodaten/Veranstaltungen/index.htm>.

Entsprechend den Empfehlungen der KVI werden diese Serviceangebote fortgeführt und für neue Scientific Use Files weiter ausgebaut. Noch stärker als bei anderen Daten, die im Rahmen der Sekundäranalyse von Forschern genutzt werden können, stellt sich bei amtlichen Mikrodaten ein erhöhter Bedarf an Datendokumentationen, da sich die dort verwendeten Definitionen und Konzepte häufig von den durch die Sozialforschung selbst erzeugten Daten unterscheiden. ZUMA stellt schon seit längerem in Kooperation mit dem Statistischen Bundesamt umfassende Datendokumentationen im WWW zur Verfügung. Durch den Aufbau des Datenbankbasierten Mikrodaten Informationssystems (MISSY, siehe 3.1) soll dieser Service erheblich verbessert werden.

2.2 Forschung und Entwicklung für besseren Service

Voraussetzung für die hohe Qualität der wissenschaftlichen Dienstleistungen sind eigenständige Forschungs- und Entwicklungsarbeiten auf Basis der betreuten Mikrodaten: Sie vermitteln ein allgemeines Hintergrundverständnis und sind unabdingbar für die Theorie- und Methodenkompetenz im Sinne des KVI-Gutachtens. Gleichzeitig fließen ihre Ergebnisse direkt in den wissenschaftlichen Service ein. Die Forschungsarbeiten tragen in Form von Mikrodaten-Tools zur Erweiterung des Analysepotenzials der Mikrodaten bei oder sind, wie beispielsweise die geplanten methodischen Untersuchungen zum Mikrozensus-Panel, mit der Erschließung neuer Scientific Use Files verbunden.

Die laufenden Forschungsarbeiten mit dem Mikrozensus stehen im Kontext der Entwicklung der Sozialstruktur und der sozialen Ungleichheit und nehmen Bezug auf die gegenwärtige inhaltliche Diskussion in diesem Teilbereich der sozialwissenschaftlichen Forschung. Die Themen der laufenden und geplanten Forschungsarbeiten sind zum Beispiel: Entwicklung und Determinanten der ethnischen Ungleichheit auf dem Arbeitsmarkt; Bildungspartizipation und soziale Differenzierung zwischen Haushalten.

Bei den Forschungsarbeiten auf Basis der EVS steht die Frage nach der sozialen Einbettung des privaten Konsums im Vordergrund. Es wird vor allem der soziale Wandel im gesellschaftlichen Stellenwert des Konsums untersucht, wie er sich z.B. in der Veränderung der Einkommensabhängigkeit von Ausgaben für unterschiedliche Konsumgüter niederschlägt. Die Methodenforschung wird sich auf Fragen der Vergleichbarmachung von Einkommens- und Verbrauchsangaben der EVS ab 1962/63 sowie auf Aspekte der Datenqualität, insbesondere fehlende Werte konzentrieren.

3. Laufende und beantragte Projekte für Mikrodaten bei ZUMA

Zum Ausbau der Dateninfrastruktur und zur Stärkung des Service, der Beratungs- und Forschungsleistungen hat der Gründungsausschuss von ZUMA beantragte Projekte zur

Förderung durch das BMBF empfohlen. Mit dem Start der Projekte MISSY, Mikrozensus-Rückerschließung und Mikrozensus-Panel wird ab Mitte 2003 gerechnet.

3.1 Mikrodaten Informationssystem - MISSY

MISSY steht für Mikrodaten InformationSSYstem. Es handelt sich um ein edv-gestütztes Datenbanksystem, welches Informationen zu sozialwissenschaftlich relevanten Mikrodaten erfasst und der Scientific Community online zur Verfügung stellen wird. MISSY ist zunächst auf den Mikrozensus beschränkt. Die Konzeption des Systems ist jedoch offen für die Integration weiterer Datenquellen.

Eine Systematisierung der datenrelevanten Informationen zum Mikrozensus durch MISSY ist einerseits notwendig, weil diese Informationen bisher häufig nur in gedruckter Form vorliegen; dem Einzelforscher in der Regel nur teilweise zugänglich sind; teilweise gar nicht vorhanden sind und vor allem unverknüpft vorliegen. Andererseits unterscheiden sich die bislang zugänglichen sieben Mikrozensus-Scientific Use Files der Zeitpunkte 1989 bis 1998 nicht nur in Bezug auf die jeweils zugrunde liegenden Erhebungsprogramme, sondern auch hinsichtlich der Variablen infolge unterschiedlicher Vergrößerungen bei der Anonymisierung. Für die Datennutzer sind deshalb vergleichende Analysen erst nach aufwändigen Vorarbeiten möglich.

MISSY stellt datenrelevante Informationen, die für die Analyse von amtlichen Mikrodaten unabdingbar sind, in systematischer Weise online zur Verfügung. Dadurch wird die Verwendung von Mikrodaten für empirisch arbeitende Wirtschafts- und Sozialforscher erheblich erleichtert und die Erschließung von amtlichen Mikrodaten nachhaltig unterstützt. Fünf Teilaspekte stehen dabei im Vordergrund: Sammlung und Zusammenführung von datenrelevanten Informationen; Erweiterung der Zugangsmöglichkeiten; effiziente Nutzung der Daten mehrerer Erhebungszeitpunkte; Steigerung der Effektivität bei der Informationsabfrage; Dokumentation und Archivierung älterer Daten.

3.2 Mikrozensus-Regionalfile

Aktuell steht der Mikrozensus nur in Form des sogenannten Grundfiles zur Verfügung, das durch differenzierte fachliche Gliederung bei einer zugleich sehr groben regionalen Gliederung gekennzeichnet ist. Diese Daten enthalten beispielsweise tief gegliederte Angaben zum Beruf oder zum Wirtschaftszweig, jedoch als Regionalangaben nur das Bundesland und eine vergrößerte Variable zur Gemeindegrößenklasse. Für die Analyse regionaler Disparitäten oder unterschiedlicher Lagen auf regionalen Arbeitsmärkten ist das Grundfile daher nicht geeignet. Die Forschung hat in den letzten Jahren zunehmend den Bedarf an einer regionalen Auswertung des Mikrozensus artikuliert (vgl. Falk/Steiner 2001: 38). Deshalb hat ZUMA in Kooperation mit der amtlichen Statistik und dem

Bundesamt für Bauwesen und Raumordnung (BBR) basierend auf den Empfehlungen des Anonymisierungsprojektes (Müller et al. 1991) ein Konzept für die Operationalisierung eines Mikrozensus-Regionalfiles entwickelt (Wirth 2000) und mit Regionalforschern diskutiert. Als regionale Einheiten werden Mikrozensus-Kreisregionen mit wenigstens 100.000 Einwohnern in der Population vorgeschlagen. Um dem Datenschutz Rechnung zu tragen und zugleich regionale Unterschiede weitestgehend zu erhalten, zum Beispiel regionale Disparitäten in Bezug auf Qualifikation und Arbeitsmarktchancen, sind Regionenspezifische Vergrößerungen der fachlichen Merkmale vorgesehen. Das Anonymisierungskonzept wurde von den in der Arbeitsgruppe vertretenen Fachstatistikern und Regionalforschern sehr positiv aufgenommen, jedoch noch nicht von den statistischen Ämtern verabschiedet. Gegenwärtig sind noch Fragen offen, die neben dem Stichprobenfehler die Veröffentlichung von Ergebnissen auf regionaler Ebene betreffen. ZUMA wird sich zusammen mit den statistischen Ämtern dafür einsetzen, dass diese Fragen bald möglichst geklärt werden, so dass Regionaldaten des Mikrozensus als Scientific Use File bereitgestellt werden können.

3.3 Mikrozensus-Panelfile

Der Mikrozensus ist als rotierende Panelstichprobe angelegt, bei der die Haushalte eines Auswahlbezirkes vier Jahre lang befragt werden, wobei jedes Jahr ein Viertel der Auswahlbezirke ausgetauscht wird. Allerdings werden wegziehende Personen und Haushalte nicht weiter befragt, sondern durch die nachziehenden Personen bzw. Haushalte ersetzt (Prinzip der Flächenstichprobe). Seit 1996 können die statistischen Ämter wieder die für die Zusammenführung benötigten Ordnungsnummern speichern und haben mit der Erstellung eines Panelfiles begonnen (Heidenreich 2002; Zühlke 2001). Aufgrund der Stichprobengröße des Mikrozensus-Panels 1996-1999, das rund 110.000 Personen umfasst, eröffnet die Weitergabe als Scientific Use File eine Vielzahl neuer inhaltlicher und methodischer Analysemöglichkeiten und kann dazu beitragen, das unter anderem von der KVI für Deutschland festgestellte Defizit an Längsschnittdaten zu verringern.

Die Voraussetzungen für die Erschließung der Daten sollen durch ein Projekt geschaffen werden, das ein zuverlässiges Verfahren der Zusammenführung der Längsschnittangaben erarbeitet, mögliche Verzerrungen aufgrund der räumlichen Mobilität für spezifische inhaltliche Fragestellungen untersucht und für deren Korrektur Gewichtung- und Hochrechnungsverfahren entwickelt und validiert sowie dokumentiert, so dass nach der Weitergabe von Scientific Use Files die Paneldaten sachgerecht verwendet werden. Da die für die Bereitstellung von Mikrozensusdaten vorliegenden Anonymisierungskonzepte nur auf Querschnittdaten bezogen sind, ist ein Anonymisierungs- und Filekonzept für die Paneldaten notwendig. Diese Ziele werden in dem gemeinsam von statistischen Ämtern und For-

schung beantragten Projekt „Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe (MZ-Panel)“ verfolgt,¹⁶ in dem ZUMA für ausgewählte Fragestellungen Selektivitätsanalysen durchführen und in Zusammenarbeit mit den statistischen Ämtern ein Anonymisierungs- und Filekonzept für das Mikrozensus-Panel entwickeln wird.

3.4 Bereitstellung älterer Mikrodaten

Der Gründungsausschuss des Rates für Sozial- und Wirtschaftsdaten hat in Anlehnung an das KVI-Gutachten (KVI 2001: 31, 295) auch die Empfehlung an das BMBF ausgesprochen, mit besonderer Dringlichkeit die Rückerschließung und Aufbereitung älterer Mikrodaten zu bearbeiten und dafür auch eigene Finanzmittel bereitzustellen. Dies betrifft beim Mikrozensus den Zeitraum vor 1989, für den noch keine Daten in Form von Scientific Use Files zur Verfügung stehen. Bei der EVS, für die Scientific Use Files der Erhebungszeitpunkte 1993 und 1998 verfügbar sind, gibt es Überlegungen, die in einem Projekt vorhandenen Erhebungen von 1962/63 bis 1988 für die Forschung zugänglich zu machen.

Der Bedarf der Wissenschaft an den jeweils *aktuellen* amtlichen *Individualdaten* ist in den letzten Jahren vielfach begründet worden und deren Notwendigkeit steht mittlerweile außer Frage. Die Frage nach dem Bedarf an Individualdaten für *zurückliegende* Zeitpunkten ist eine ebenso fast rhetorisch anmutende Frage, wie die Frage nach der Notwendigkeit von aktuellen Daten (Müller/Hauser 1987: 63). Eine der zentralen Fragestellungen in den Sozial- und Wirtschaftswissenschaften ist die Untersuchung des wirtschaftlichen- und gesellschaftlichen Wandels. Dass der Mikrozensus und die EVS hierfür einzigartige Datengrundlagen bilden, ist evident.

3.4.1 Mikrozensus

Aufgrund der kontinuierlichen Erhebung des Mikrozensus seit 1957 und eines in den zentralen Bereichen weitgehend kontinuierlich erhobenen Fragenprogramms – trotz aller Probleme aufgrund von nicht zu vermeidenden Brüchen oder Änderungen in einem über 40jährigen Erhebungsprogramm – sind die Mikrozensusdaten eine wertvolle Quelle zur langfristigen Analyse des gesellschaftlichen und wirtschaftlichen Wandels im Haushalts- und Familienkontext mittels Kohorten- und Längsschnittanalysen. Zudem deckt der Mikrozensus aufgrund seiner Konzeption als Mehrzweckstichprobe mehrere zentrale gesellschaftliche Themenbereiche ab.¹⁷ Der Erhebungszeitraum, aus dem einzelne Erhe-

¹⁶ Projektpartner sind das Statistische Bundesamt, das Landesamt für Datenverarbeitung und Statistik NRW, die Universität Frankfurt (Prof. Dr. U. Rendtel) und ZUMA.

¹⁷ Die in der empirischen Sozialforschung verfügbaren Datensätze, mit denen Längsschnittanalysen möglich sind, reichen zwar z.T. auch in die achtziger Jahre zurück (ALLBUS ab 1980; SOEP ab 1984; Wohlfahrtssurvey ab 1978), haben aber einige Einschränkungen gegenüber dem Mikrozensus

bungszeitpunkte ausgewählt werden sollen, umfasst den Zeitraum zwischen 1957 (Mikrozensus erstmals erhoben) und 1989 (1. Scientific Use File). In einem Teilprojekt bei ZUMA wird die Frage bearbeitet, welche konkreten Zeitpunkte ausgewählt werden sollen und welche Anonymisierungsmaßnahmen ergriffen werden müssen. Nach der Bereitstellung der Datenfiles durch die statistischen Ämter werden zudem, wie auch bei den neueren Mikrozensen, umfangreiche Dokumentationsmaterialien zu den Daten und Tools zur Lösung von Fragen der Vergleichbarkeit bereitgestellt.

3.4.2 Einkommens- und Verbrauchsstichprobe

Die seit 1962/63 im fünfjährigen Turnus durchgeführte EVS ist eine wichtige Datenbasis, wenn nach sozio-ökonomischen Gruppen differenzierte Gesamtdarstellungen der Markteinkommenskomponenten, der empfangenen Sozialleistungen, der persönlichen Steuern und Sozialabgaben, der Ersparnis sowie der Strukturen und Entwicklungen des Verbrauchs der Privathaushalte im Zeitverlauf untersucht werden sollen. Neben demographischen Angaben zur sozio-ökonomischen Zusammensetzung der Haushalte bzw. zu den Haushaltsmitglieder werden auch noch wichtige Elemente des Vermögens und der Schulden sowie die Ausstattung mit technischen Gebrauchsgütern erfragt. Insbesondere liegt die Stärke der EVS in der Verbindung von sozialer und ökonomischer Lage einerseits und Konsummustern andererseits.¹⁸

Die Variablenauszüge der EVS-Daten des ehemaligen Sonderforschungsbereichs 3 der Universitäten Frankfurt und Mannheim sind ab der ersten Erhebung 1962/63 bis einschließlich 1988 in einer Datenbank des EVS-Projekts von Prof. Richard Hauser an der Universität Frankfurt gespeichert.¹⁹ Die Daten der einzelnen Erhebungszeitpunkte umfassen zwischen rund 33.000 bis 46.000 Haushalte. Aufgrund der Weitergabe der Files ab 1973 als „absolut anonymisierte Daten“ sind verschiedene Angaben jedoch stark vergrößert und in ihren Analysemöglichkeiten begrenzt. Um das Analysepotenzial dieser zeitlich weit zurück reichenden Datenbestände für die Forschung zu erschließen, ist geplant die Daten des EVS-Projekts an das Forschungsdatenzentrum des Statistischen Bundesamtes und ZUMA zu

(Stichprobengröße, Ausschöpfung etc.). Zu den Vorteilen aber auch den Einschränkungen des Mikrozensus gegenüber Daten der Umfrageforschung siehe Lüttinger/Riede 1997; Schimml-Neimanns 2002.

¹⁸ Hinsichtlich der Generalisierbarkeit der Ergebnisse ergeben sich zwar Einschränkungen mit Bezug auf die Durchführung als Quotenstichprobe, für die im allgemeinen eine weitgehend unbekannte Selektivität der Teilnehmer vermutet wird, sowie hinsichtlich der Untererfassung oberer Einkommensgruppen, jedoch gibt es keine alternative Datenbasis, mit der diese Fragestellungen des sozialen Wandels ab Anfang der 60er-Jahre untersucht werden könnten.

¹⁹ Die Einzeldaten der EVS 1962/63 und 1969 sind nur noch in diesem Projekt und selbst nicht mehr im Statistischen Bundesamt vorhanden (siehe zu Datenbeschreibungen und Projektergebnissen <http://www.wiwi.uni-frankfurt.de/professoren/hauser/evspapiere.htm>).

übermitteln. Nach Bereitstellung der Daten durch das Forschungsdatenzentrum des Statistischen Bundesamtes als Scientific Use Files wird ZUMA Dokumentationsmaterialien und Tools zur Vergleichbarkeit der Daten im Zeitverlauf anbieten.

4. Der neue Verbund: Forschungsdaten- und Servicezentren

In der Einrichtung von Forschungsdaten- und Servicezentren, die den Zugang der Wissenschaft zu Daten der amtlichen Statistik und anderer Datenproduzenten im öffentlichen Bereich erleichtern sollen und in der Ausweitung der kostengünstigen Bereitstellung von anonymisierten Mikrodaten für die Wissenschaft, sieht die KVI wesentliche Aufgaben zur Verbesserung der Dateninfrastruktur. Die hierzu notwendigen Arbeiten und Dienstleistungen werden von folgenden Forschungsdaten- und Servicezentren angeboten:

Auf der Seite der Datenproduzenten: Forschungsdatenzentren des Statistischen Bundesamtes und der Statistischen Landesämter sowie das Forschungsdatenzentrum der Bundesanstalt für Arbeit. Auf der Seite der Wissenschaft: neben ZUMA noch das Internationale Datenservicezentrum des Instituts zur Zukunft der Arbeit (IZA) in Bonn. Im Folgenden werden das Servicezentrum am Institut zur Zukunft der Arbeit und die Forschungsdatenzentren kurz vorgestellt.

4.1 Internationales Datenservicezentrum am Forschungsinstitut zur Zukunft der Arbeit (IZA)

Im Vordergrund der Dienstleistung steht der Nachweis und die Vermittlung zu arbeitsmarktrelevanten Daten. Als Metainformationsportal sollen deutsche Dateninformationen an ausländische Forscher sowie ausländische Dateninformationen an deutsche Wissenschaftler vermittelt werden.

Auf der Grundlage bestehender Thesauri soll in Kooperation mit den übrigen Forschungsdaten- und Servicezentren ein Modul für den Bereich Arbeitsmarktdaten entwickelt werden. Dieser Thesaurus und weitere Metainformationen werden Daten unter arbeitsmarktökonomischer Perspektive dokumentieren und erschließen. Das Internationale Datenservicezentrum stellt damit eine wichtige Ergänzung von Serviceleistungen bestehender Einrichtungen dar.

Soweit rechtlich möglich, soll die Datennutzung im Datenservicezentrum erfolgen. Nach Maßgabe rechtlicher Datennutzungsbestimmungen sollen Gastwissenschaftler des IZA im Datenservicezentrum Auswertungsmöglichkeiten erhalten. Der Aufbau einer kontrollierten Datenfernverarbeitung erfolgt unter Einhaltung der rechtlichen Nutzungsbestimmungen.

Kontakt:

*Georgios Tassoukis
Forschungsinstitut zur Zukunft der Arbeit
Postfach 7240
53072 Bonn
Telefon: 0228-3894-503
email: tassoukis@iza.org*

4.2 Forschungsdatenzentrum des Statistischen Bundesamtes

Im Forschungsdatenzentrum des Statistischen Bundesamtes, mit Standorten in Wiesbaden und Bonn, das im Oktober 2001 seine Arbeit aufgenommen hat, sollen Daten zu den gesellschaftlichen, ökonomischen, ökologischen und kulturellen Verhältnissen und Entwicklungen in Deutschland unter Wahrung der statistischen Geheimhaltung für die deutsche und internationale Wissenschaft nutzbar gemacht werden. Das Forschungsdatenzentrum wird sich folgenden Kernaufgaben widmen (Zwick 2003):

- Klärung von Grundsatzfragen des Datenzugangs unter Beachtung der rechtlichen Anforderungen des Datenschutzes und des Statistikgeheimnisses.
- Bereitstellung und Dokumentation von wichtigen Datensätzen als Scientific Use Files und Public Use Files.
- Aufbau einer Infrastruktur für kontrollierte Datenfernverarbeitung, langfristig auch durch den Einsatz von Internetverbindungen.
- Ermöglichung von Gastaufenthalten von externen Wissenschaftlern.

Kontakt:

*Markus Zwick
Statistisches Bundesamt
Institut Ib / Forschungsdatenzentrum
65180 Wiesbaden
Telefon: 0611-75-4220
email: markus.zwick@destatis.de*

4.3 Forschungsdatenzentrum der Statistischen Landesämter

Die Zielsetzung und das Angebot des Forschungsdatenzentrums der Statistischen Landesämter, das mit Standorten in 16 Bundesländern seit April 2002 seine Arbeit aufgenommen hat, entspricht der des Forschungsdatenzentrums des Statistischen Bundesamtes: den Zugang zu amtlichen Daten zu erleichtern, flexiblere und intensivere Datenanalysen zu ermöglichen, Beratung, Metainformationen und Service zu den im Forschungsdatenzentrum nutzbaren Daten anzubieten sowie diese Daten durch Gastwissenschaftler-Arbeitsplätze und kontrollierte Datenfernverarbeitung zugänglich zu machen (Bauer 2003; Bech-

told/Zühlke 2002). Hervorzuheben ist, dass aufgrund der föderalen Struktur der amtlichen Statistik über 90 Prozent aller amtlichen Mikrodaten bei den Statistischen Landesämtern vorliegen, welche bislang jedoch lediglich in den jeweiligen Landesämtern zur Verfügung standen. Ein wesentliches Ziel des Forschungsdatenzentrums der Statistischen Landesämter ist es daher, diese Daten länderübergreifend für die wissenschaftliche Nutzung zugänglich zu machen. Hierfür soll eine organisatorische und technische Infrastruktur geschaffen werden, welche die Nutzung der betreffenden Mikrodaten aller Landesämter an jedem der Standorte des Forschungsdatenzentrums ermöglicht. Die Ausgestaltung des Datenangebotes orientiert sich dabei eng an den Präferenzen der Wissenschaft, welche im vergangenen Jahr mit einer Nutzerbefragung evaluiert wurden (Zühlke/Hetke 2003).

Kontakt:

Dr. Sylvia Zühlke

Geschäftsstelle des Forschungsdatenzentrums der Statistischen Landesämter

c/o Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen

Postfach 10 11 05

40002 Düsseldorf

Telefon: 0211-9449-2547

email: sylvia.zuehlke@lds.nrw.de

4.4 Forschungsdatenzentrum der Bundesanstalt für Arbeit

Die Bundesanstalt für Arbeit zählt neben den statistischen Ämtern zu den wichtigsten Datenproduzenten in Deutschland. Insbesondere liegen umfangreiche prozessproduzierte Mikrodaten über den Arbeitsmarkt vor, die im Verwaltungsablauf regelmäßig anfallen. Sie sind aufgrund ihrer Vollständigkeit und - nach entsprechender Aufbereitung - als ereignisorientierte Daten mit Möglichkeiten der Verknüpfung von Personen- und Betriebsebene für die Arbeitsmarktforschung von außerordentlicher Bedeutung. Hauptziel des Forschungsdatenzentrums der Bundesanstalt für Arbeit ist die Aufbereitung und Bereitstellung von solchen Einzeldaten der Bundesanstalt für Arbeit, die nicht ohne erheblichen Informationsverlust anonymisiert und als Scientific Use Files an die Forschung weitergegeben werden können (Bender 2002). Externe Wissenschaftler können als Gastwissenschaftler mit entsprechender datenschutzrechtlicher Genehmigung in einem abgeschotteten Bereich die schwach anonymisierten Mikrodaten (z.B. Daten der Beschäftigtenstatistik, Daten von Teilnehmern an arbeitsfördernden Maßnahmen, Linked-Employer-Employee-Daten) analysieren oder mittels Programmerstellung über Datenfernverarbeitung nach dem Modell der für das IAB-Betriebspanel eingerichteten Schalterstelle nach ihren Vorgaben auswerten lassen. Hierfür sollen Datendokumentationen (Metadaten) und anonymisierte Testdatensätze umfassend

aufbereitet und integriert werden. Daneben soll die Bereitstellung von Scientific Use Files zur Weitergabe an externe Forscher fortgeführt und verstärkt werden.

Kontakt:

Dr. Jürgen Passenberger

Bundesanstalt für Arbeit, Institut für Arbeitsmarkt- und Berufsforschung

Arbeitsbereich 9

Regensburger Str. 104

90327 Nürnberg

Telefon: 0911-179-3077

email: Juergen.Passenberger@iab.de

5. Ausblick

Die KVI hat in ihrem Gutachten die herausragende Bedeutung von amtlichen Mikrodaten für die empirische Sozial- und Wirtschaftswissenschaft betont und eine Reihe von Möglichkeiten zur Verbesserung der Dateninfrastruktur festgestellt. In Bezug auf den Zugang zu Mikrodaten, auf den sich dieser Überblick konzentriert, sind mit den vom Gründungsausschuss zur Förderung durch das BMBF empfohlenen Vorhaben die Grundsteine dafür gelegt worden, dass auch in Deutschland eine für die Arbeit mit amtlichen Mikrodaten notwendige Infrastruktur entstehen und der Rückstand im Vergleich zu vielen anderen Ländern aufgeholt werden kann (vgl. Mayer 2003). Am Schluss dieser Übersicht ist aber auch darauf hinzuweisen, dass noch viel für eine Konsolidierung zu tun ist.

Die nach internationalen Vorbildern neu geschaffenen Forschungsdatenzentren bei den Datenproduzenten sollen für die Forschung Möglichkeiten in Form von Gastwissenschaftler-Modellen oder Datenfernauswertung eröffnen, damit auch Analysen von Daten durchgeführt werden können, die nicht als Scientific Use Files weitergegeben werden können.²⁰

Aus Sicht der empirischen Sozial- und Wirtschaftsforschung sind Scientific Use Files aufgrund „(...) der Flexibilität der Nutzung, der durch Anonymisierung relativen einfachen Sicherung des Datenschutzes und der mit dieser Lösung verbundenen Kostenvorteile (...)“ (KVI 2001: 270) mit höchster Priorität zu entwickeln und weiterzugeben. Ergänzend sollten, wo möglich, öffentlich zugängliche Datensätze (Public Use Files) bereitgestellt werden. Dies wurde nicht nur durch eine Umfrage des Forschungsdatenzentrums der Statistischen

²⁰ Um gegenwärtig bestehende Nutzungsrestriktionen (vgl. KVI 2001: 170-172) aufzuheben und eine den Bedürfnissen der Forschung entgegenkommende Arbeitsweise zu ermöglichen, sind Änderungen des Bundesstatistikgesetzes notwendig.

Landesämter (Zühlke/Hetke 2002) unter Sozial- und Wirtschaftswissenschaftlichen Instituten, sondern auch auf der Konferenz für Sozial- und Wirtschaftsdaten betont.²¹

Hinsichtlich amtlicher Mikrodaten wird ZUMA weiterhin Verbesserungen des Zugangs und der möglichst guten Nutzbarkeit dieser Daten für die Forschung unterstützen. Neben der inhaltlichen und zeitlichen Erweiterung beim Mikrozensus (ältere Files, Regionalfiler, Panelfiler) und der EVS bringt ZUMA darüber hinaus seine langjährige Expertise zur Umsetzung faktischer Anonymität in Projekte zur Erschließung neuer Datenbasen, auch solchen anderer Datenproduzenten, ein. Der bereits aufgebaute wissenschaftliche Service zu den Scientific Use Files des Mikrozensus und der EVS wird erweitert. Insbesondere durch das MISSY-Projekt soll der inhaltliche wie methodische Zugang zum Mikrozensus als Datenquelle für sozialwissenschaftliche Untersuchungen erheblich erleichtert werden.

Wie bei den Forschungsdatenzentren in Bezug auf die Nutzungsrechte von Forschern sind aber auch bei Scientific Use Files noch Restriktionen festzustellen. Diese Files können aus rechtlichen Gründen gegenwärtig nicht an Forscher ins Ausland weitergegeben werden. Damit international vergleichende Analysen nicht mehr erschwert und Kooperationen mit Forschern im Ausland erleichtert werden, sind Anpassungen des Bundesstatistikgesetzes an das „Statistikgesetz“ der EU (Verordnung (EG) Nr. 322/97 vom 17.2.1997) dringend notwendig.²²

Wie oben erwähnt, ist mit finanzieller Unterstützung des BMBF die kostengünstige Bereitstellung ausgewählter Mikrodaten des Mikrozensus, der EVS, der deutschen Stichprobe des ECHP und der Zeitbudgeterhebung entfristet worden. Insbesondere vor dem Hintergrund, dass vor allem für Forschungsarbeiten im Umfeld der Politikberatung dringend aktuelle Daten benötigt werden, ist es sehr wichtig, dass - ebenfalls mit Unterstützung des BMBF - die statistischen Ämter und die Bundesanstalt für Arbeit aktuelle Daten kostengünstig als Scientific Use Files bereitstellen werden. Gerade die Bereitstellung zeitnaher Scientific Use Files kann die Bedeutung amtlicher Mikrodaten für die Gesellschaftsanalyse belegen und damit zugleich zeigen, dass die zurzeit nur projektmäßig geförderte, im Aus- und Aufbau begriffene Infrastruktur und Datenweitergabep Praxis dauerhaft gesichert und konsolidiert werden muss.

²¹ Siehe die Pressemitteilung des Gründungsausschusses <http://www.destatis.de/download/d/veranst/presse.doc>.

²² Zu internationalen Aspekten des Datenzugangs siehe auch den Abschlussbericht der OECD Follow Up Group on Issues of Access to Publicly Funded Research Data (2003).

Korrespondenzadresse

Dr. Paul Lüttinger
ZUMA
Postfach 12 21 55
68 072 Mannheim
email: luettinger@zuma-mannheim.de

Literatur

- Bauer, P., 2003: Materialien zur 1. Konferenz für Sozial- und Wirtschaftsdaten. <http://www.dstatg.de/archiv/KVI_KSWD2003_bauer.pdf>.
- Bechtold, S./Zühlke, S., 2002: Researchers, come in, please. *Sigma* 2: 16-17.
- Bender, S., 2002: Forschungsdatenzentrum in der BA. S. 16 in: IAB Werkstattbericht Nr. 2/20.3.2002. Nürnberg: IAB <<http://doku.iab.de/werkber/2002/wb0202.pdf>>.
- Hauser, R./Wagner, G./Zimmermann, K.F., 1998: Erfolgsbedingungen empirischer Wirtschaftsforschung und empirisch gestützter wirtschafts- und sozialpolitischer Beratung. Ein Memorandum. *Allgemeines Statistisches Archiv* 82: 369-379.
- Falk, M./Steiner, V., 2001: Expertise für die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik im Bereich Erwerbstätigkeit und Einkommen. In: Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) (Hg.): Wege zur besseren informationellen Infrastruktur. Baden-Baden: Nomos. [CD-ROM Beilage zur Buchausgabe]
- Heidenreich, H.-J., 2002: Längsschnittdaten aus dem Mikrozensus: Basis für neue Analysemöglichkeiten. *Allgemeines Statistisches Archiv* 86(2): 213-231.
- Köhler, S./Schimpl-Neimanns, B./Schwarz, N., 2000: Pilotprojekt zur Erleichterung der Nutzungsmöglichkeiten von faktisch anonymisierten Mikrodaten. *Wirtschaft & Statistik*: 30-37.
- KVI (Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik) (Hrsg.), 2001: Wege zu einer besseren informationellen Infrastruktur. Gutachten der vom Bundesministerium für Bildung und Forschung eingesetzten Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik. Baden-Baden: Nomos.
- Lüttinger, P./Riede, T., 1997: Der Mikrozensus: Amtliche Daten für die Sozialforschung. *ZUMA-Nachrichten* 41: 19-43.
- Lüttinger, P., 1999: Sozialstrukturanalysen mit dem Mikrozensus. *ZUMA-Nachrichten Spezial* Band 6. ZUMA, Mannheim.
- Mayer, K.U., 2003: Materialien zur 1. Konferenz für Sozial- und Wirtschaftsdaten. <http://www.dstatg.de/archiv/KVI_KSWD2003_mayer.pdf>.

- Müller, W., 2001: Konsequenzen für die Methodenausbildung aus dem Gutachten der Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI). ZUMA-Nachrichten 49: 81-99.
- Müller, W., 1999: Der Mikrozensus als Datenquelle sozialwissenschaftlicher Forschung. S. 7-27 in: Lüttinger, P. (Hrsg.): Sozialstrukturanalysen mit dem Mikrozensus. ZUMA-Nachrichten Spezial Band 6. ZUMA, Mannheim.
- Müller, W./Blien, U./Knoche, P./Wirth, H., 1991: Die faktische Anonymität von Mikrodaten. Forum der Bundesstatistik, Band 9, hrsg. v. Statistischen Bundesamt. Stuttgart: Metzler-Poeschel.
- Müller, W./Hauser, R., 1987: Der Bedarf der Wissenschaft an anonymisierten Einzelangaben. S. 61-78 in: Statistisches Bundesamt (Hrsg.), Nutzung von anonymisierten Einzelangaben aus den Daten der amtlichen Statistik: Bedingungen und Möglichkeiten. Schriftenreihe Forum der Bundesstatistik; 5. Stuttgart: Kohlhammer.
- OECD Follow Up Group on Issues of Access to Publicly Funded Research Data, 2003: Promoting Access to Public Research Data for Scientific, Economic, and Social Development. Final Report. <http://dataaccess.ucsd.edu/Final_Report_2003.pdf>.
- Papastefanou, G., 2002: Consumer Panel Data: A ZUMA research service. S. 623-643 in: Schmollers Jahrbuch, Journal of Applied Social Science Studies, 122.
- Schimpl-Neimanns B., 2002: Anwendungen und Erfahrungen mit dem Scientific Use File des Mikrozensus. ZUMA-Arbeitsbericht Nr. 2002/01.
- Statistisches Bundesamt (Hrsg.) 1999: Kooperation zwischen Wissenschaft und amtlicher Statistik - Praxis und Perspektiven. Beiträge zum Symposium am 31. Mai/1. Juni 1999 in Wiesbaden. Band 34, Schriftenreihe Forum der Bundesstatistik. Stuttgart: Metzler-Poeschel.
- Wirth, H., 2000: Konzeption eines Mikrozensus-Regionalfiles. Expertengespräch Mikrozensus-Regionalfile. (Ms.) Mannheim: ZUMA.
- Wirth, H., 2003: Angriffsszenarien auf wirtschaftsstatistische Einzeldaten - Ein Überblick. Ms. zum Vortrag auf dem Workshop „Anonymisierung wirtschaftsstatistischer Einzeldaten“, 20./21. März 2003, Tübingen, <http://www.uni-tuebingen.de/iaw/pdf/Beitrag_Wirth.pdf>.
- Zühlke, S., 2001: Längsschnittanalysen auf der Basis des Mikrozensus. Methodische Probleme und Lösungsansätze. Statistische Analysen und Studien NRW (4): 3-13.
- Zühlke, S./Hetke, U., 2002: Datenbedarf der Wissenschaft. Ein Bericht des Forschungsdatenzentrums der statistischen Landesämter über die erste Nutzerbefragung. S. 3-10 in: Landesamt für Datenverarbeitung und Statistik Nordrhein Westfalen (Hrsg.), Statistische Analysen und Studien NRW, Band 6.
- Zühlke, S./Hetke, U., 2003: Datenbedarf und Datenzugang: Ergebnisse der ersten Nutzerbefragung des Forschungsdatenzentrums der statistischen Landesämter. Erscheint in: Allgemeines Statistisches Archiv, 3.

ZUMA-Nachrichten Nr. 39, 1996: Zur Verbesserung der Zugangsmöglichkeiten zu Mikrodaten der amtlichen Statistik, S. 171-178.

ZUMA-Nachrichten Nr. 41, 1997: Weitere amtliche Mikrodaten für die Forschung leichter nutzbar, S. 182-187.

Zwick, M., 2003: Materialien zur 1. Konferenz für Sozial- und Wirtschaftsdaten. <http://www.dstatg.de/archiv/KVI_KSWD2003_zwick.pdf>.

Anhang

Tabelle 1: Gegenwärtig verfügbare Scientific Use Files

Beschäftigtenpanel der BA	1998-2002
Deutsche Stichprobe des Europäischen Haushaltspanels	1994-1996
Einkommens- und Verbrauchsstichprobe	1993, 1998
IAB-Beschäftigtenstichprobe	1975-1995 (Basisfile), 1975-1997 (Regionalfile)
Mikrozensus	1989, 1991, 1993, 1995, 1996, 1997 und 1998. Im Lauf des Jahres 2003 kommen noch die Mikrozensus der Jahre 1999, 2000 und 2001 hinzu.
Niedrigeinkommens-Panel (NIEP)	1999-2002
Zeitbudgeterhebung	1991/92

Zu den Zugangsmöglichkeiten zu diesen Daten siehe http://www.gesis.org/Dauerbeobachtung/Mikrodaten/Informationsquellen/weitere_Mikrodaten.htm

Zum Mikrozensus und zur Einkommens- und Verbrauchsstichprobe sind eine große Anzahl von Informationen über das WWW-Angebot von ZUMA abrufbar. (<http://www.gesis.org/Dauerbeobachtung/Mikrodaten/index.htm> und <http://www.gesis.org/Dauerbeobachtung/Einkommen/index.htm>).

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 2003/01

Daniel Lehnert & Annette Kohlmann
Mikrozensus 1998. Dokumentation und Datenaufbereitung.

ZUMA-Methodenbericht 2003/02

Daniel Lehnert & Annette Kohlmann
Vergleich der Mikrozensus 1997 und 1998.

ZUMA-Methodenbericht 2003/03

Janet Harkness, Sabine Klein & Evi Scholz
ISSP Study Monitoring 1999. Report to the ISSP General Assembly on monitoring work undertaken for the ISSP by ZUMA.

FREI-RÄUME UND FREI-ZEITEN: RAUM-NUTZUNG UND ZEIT-VERWENDUNG IM GESCHLECHTERVERHÄLTNIS



Wie groß sind die FREI-Räume und FREI-Zeiten, über die Männer und Frauen in ihrem Alltag verfügen können? Wie unterschiedlich gestaltet sich die Nutzung von (öffentlichem) Raum zwischen den Geschlechtern? Wie werden öffentlicher Raum und öffentliche Zeit „erobert“, wie geplant, wie wirken sich flexibilisierte Arbeitszeiten aus und welche Zeitwänge wirken wo auf wen ganz besonders? Diesen Fragen widmen sich bisher erstmals in interdisziplinärer Perspektive Wissenschaftler/-innen aus Geographie, Soziologie, Psychologie, Stadt- und Landschaftsplanung, Raumplanung, Wirtschaftswissenschaften, Sozialpädagogik und Sportwissenschaften sowie Gleichstellungsbeauftragte, Planerinnen und Politikerinnen. Im ersten Teil des Bandes stehen die FREI-Räume im Geschlechterverhältnis im Vordergrund. Nach einem Überblick über die Entwicklung der raumbezogenen Genderforschung werden die Themenbereiche Aneignung von öffentlichem Raum im Lebenslauf, Angsträume (kritische Bemerkungen zu einem Begriff) und Planungsräume für Männer und Frauen diskutiert. Im zweiten Teil der FREI-Zeiten im Ge-

schlechterverhältnis widmen sich die Autorinnen den Themen Arbeitszeiten und ihre zunehmende Flexibilisierung – Chance oder Hindernis sowie den Zeitwängen im Alltag. Dabei werden u.a. Projekte zur Zeitplanung der Stadt Bremen sowie Erfahrungen aus Transformationsländern vorgestellt.

Dieser Band bietet allen, die sich für Stadt-/Regionalplanung in einer Gender-Perspektive, für Arbeits- und Familien-Zeit-Planung und die Verknüpfung dieser Bereiche interessieren, Einblick in die aktuelle Forschungsdiskussion.

*FREI-Räume und FREI-Zeiten: Raum-Nutzung und
Zeit-Verwendung im Geschlechterverhältnis*

Caroline Kramer (Hrsg.)

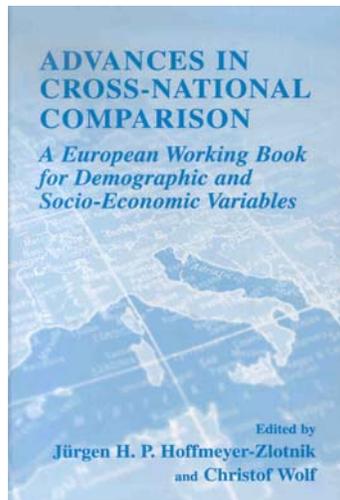
Schriftenreihe des Heidelberger Instituts für Interdisziplinäre
Frauen- und Geschlechterforschung, Band 5

Baden-Baden: Nomos Verlagsgesellschaft 2002

ISBN 3-783789-08338-0, 252 Seiten, € 24

ADVANCES IN CROSS-NATIONAL COMPARISON:

A EUROPEAN WORKING BOOK FOR DEMOGRAPHIC AND SOCIO-ECONOMIC VARIABLES



As the European Union grows together we are faced with an increasing number of European surveys and statistics that rely on national instruments or questionnaires. We lack valid instruments with functional equivalence that allow international comparative analysis. This problem is especially important in the case of demographic and socio-economic variables. To overcome these problems this book offers rules of comparison, tested Instruments and examples for the measurements of basic demographic and socio-economic variables. It is mainly written for two groups. First, researchers and practitioners involved in comparative Research in Europe. Second, researchers working with data of the statistical offices of European countries and data from institutions of the European Union.

*Advances in Cross-National Comparison:
A European Working Book for Demographic and Socio-Economic Variables*
Jürgen H.P. Hoffmeyer-Zlotnik & Christof Wolf (Eds.)
Kluwer Academic/Plenum Publishers: New York, 2003
ISBN 0-306-47731-9; 420p; \$ 150, £ 96, € 153

FÜHRUNGSINSTRUMENT MITARBEITERBEFRAGUNG



Mitarbeiterbefragungen gehören zu den am häufigsten verwendeten Instrumenten in der Organisationsführung und –entwicklung. Sie dienen dazu, den Zustand einer Organisation im Hinblick auf die “weichen Faktoren”, wie Führung und Mitarbeiter, zu messen und den strategischen Fortschritt zu beschleunigen. Das Buch stellt alle für die erfolgreiche Durchführung einer Mitarbeiterbefragung notwendigen Überlegungen und Schritte praxisorientiert, detailliert und wissenschaftlich fundiert dar.

Die dritte Auflage des Buches wurde gründlich überarbeitet, auf den neuesten Stand gebracht und erheblich erweitert. Dabei wurden die Folgeprozesse von Mitarbeiterbefragungen, beginnend mit der Ergebnispräsentation für die Geschäftsleitung bis hin zu verschiedenen Workshops auf Management- und Arbeiterebene, deutlicher in den Vordergrund gerückt und durch zahlreiche Tipps, Hinweise und Checklisten praxisnah beschrieben.

*Führungsinstrument Mitarbeiterbefragung
3. überarbeitete und erweiterte Auflage*

Ingwer Borg

Hogrefe Verlag, Göttingen, Bern, Toronto, Seattle, 2003 (2000, 1995)
ISBN 3-8017-1716-X, 458 Seiten, € 39,95 / sFr. 67,-