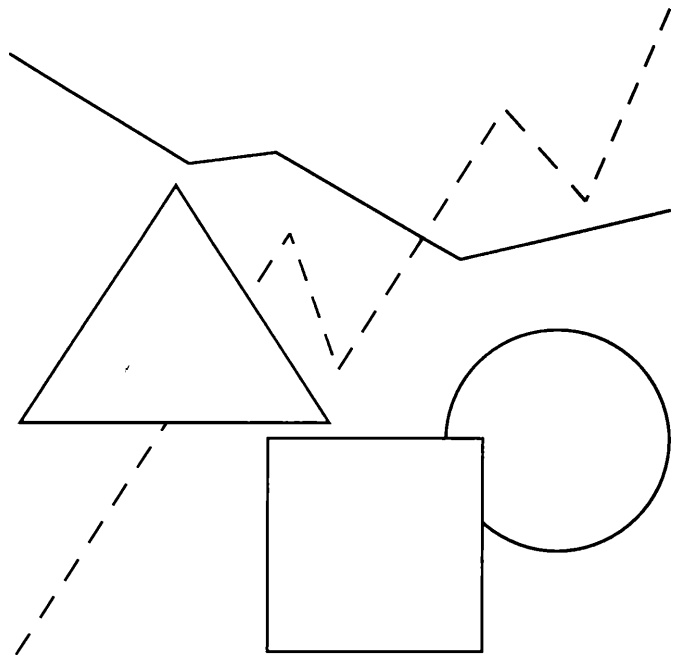


Pretest und Weiterentwicklung von Fragebogen



Band 9 der Schriftenreihe
Spektrum Bundesstatistik

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Pretest und Weiterentwicklung von Fragebogen / Statistisches

Bundesamt. – Stuttgart: Metzler-Poeschel, 1996

(Band 9 der Schriftenreihe Spektrum Bundesstatistik; Bd. 9)

ISBN 3-8246-0472-8

NE: Deutschland / Statistisches Bundesamt; Schriftenreihe Spektrum
Bundesstatistik



Weitere Informationen zu dieser Veröffentlichung können direkt beim Statistischen Bundesamt erfragt werden:

Gruppe IB, Telefon: 0611 / 75 - 20 77 oder Fax: 0611 / 72 40 00

Herausgeber: Statistisches Bundesamt, Wiesbaden



Informationen: Statistisches Bundesamt

Allgemeiner Auskunftsdienst

65180 Wiesbaden

- Telefon: 0611 / 75 - 24 05
- Telefax: 0611 / 75 33 30
- T-Online (Btx): * 4 84 84#
- Internet: <http://www.statistik-bund.de>

Zweigstelle Berlin

Postfach 276, 10124 Berlin

- Telefon: 030 / 23 24 68 66
- Telefax: 030 / 23 24 68 72

Verlag: Metzler-Poeschel, Stuttgart

Verlagsauslieferung: Hermann Leins GmbH & Co. KG

Postfach 11 52

72125 Kusterdingen

Telefon: 07071 / 93 53 50

Telefax: 07071 / 3 36 53

Erschienen im Mai 1996

Preis: DM 17,80

Bestellnummer: 1030509 - 96900

ISBN: 3-8246-0472-8

© Statistisches Bundesamt, Wiesbaden 1996

Alle Rechte vorbehalten. Es ist insbesondere nicht gestattet, ohne ausdrückliche Genehmigung des Statistischen Bundesamtes diese Veröffentlichung oder Teile daraus für gewerbliche Zwecke zu übersetzen, zu vervielfältigen, auf Mikrofilm/-fiche zu verfilmen oder in elektronische Systeme einzuspeichern.

Recyclingpapier aus 100 % Altpapier.

Vorwort

Die wissenschaftliche Tagung „Pretest und Weiterentwicklung von Fragebogen“, deren Tagungsband hiermit vorgelegt wird, hat am 22. und 23. Juni 1995 im Gerhard-Fürst-Saal des Statistischen Bundesamtes in Wiesbaden stattgefunden. Sie wurde gemeinsam vom Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V. (ADM), der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e. V. (ASI) und dem Statistischen Bundesamt veranstaltet.

Die Veranstaltung bot nicht nur Referate zu einzelnen Facetten von Pretests auf hohem wissenschaftlichem Niveau, sondern auch die Möglichkeit zu einem anregenden und fruchtbaren Erfahrungsaustausch zwischen Marktforschern, Sozialwissenschaftlern und amtlichen Statistikern über einen gemeinsamen Teilaspekt ihrer täglichen Arbeit.

Dank gilt den Referentinnen und Referenten sowie allen, die durch ihre Diskussionsbeiträge die Veranstaltung bereichert haben. Ebenfalls zu danken ist denen, die mit ihrem Engagement bei der Vorbereitung und Durchführung zum Gelingen der Veranstaltung beigetragen haben. Herr Prof. Dr. Mohler vom Zentrum für Umfragen, Methoden und Analysen (ZUMA) in Mannheim hat die Tagung moderiert. Für seinen Einsatz gebührt ihm ein ganz besonderer Dank.

Der vorliegende Band der Reihe „Spektrum Bundesstatistik“ geht über einen reinen Tagungsband hinaus. Er enthält neben den Tagungsbeiträgen des Moderators und der Referentinnen und Referenten auch zwei Erfahrungsberichte über die Pretests zu zwei Erhebungen nach § 7 Bundesstatistikgesetz. Dabei handelt es sich um die Beiträge *Pretest zur Erhebung „Neukonzeption der laufenden Wirtschaftsrechnungen“* und *Fragebogen-Pretest zur „Erfassung von Produktionstagen und Betriebszeiten“*, die beide im Auftrag des Statistischen Bundesamtes von ZUMA durchgeführt wurden. Die Zusammenarbeit bei diesen Pretests zeigt, daß die Kooperation zwischen Instituten und Statistischem Bundesamt über den reinen theoretischen und informellen Austausch hinausgeht und zu konkreten, für alle Beteiligten praktisch verwertbaren Ergebnissen führen kann.

Wiesbaden, im April 1996

Der Präsident des Statistischen Bundesamtes

Johann Hahlen

Inhalt

	Seite
Vorwort	3
 <i>Günter Kopsch</i>	
Begrüßung	5
 <i>Peter Ph. Mohler / Rolf Porst</i>	
Pretest und Weiterentwicklung von Fragebogen – Einführung in das Thema	7
 <i>Erwin K. Scheuch</i>	
Die Notwendigkeit von Pretests zur Vorbereitung statistischer Erhebungen	16
 <i>Horst Becker</i>	
Der klassische Pretest und seine Techniken	28
 <i>Klaus Haupt</i>	
Bewertung und Beurteilung von Fragebogen durch Gruppendiskussionen	35
 <i>Michaela Wänke</i>	
Kognitionspsychologische Ansätze zum Pretest von Fragebogen	46
 <i>Elisabeth Noelle-Neumann</i>	
Die Fragebogenkonferenz	55
 <i>Hans Akkerboom</i>	
Labor für die Entwicklung und den Test von Erhebungsinstrumenten	66
 <i>Manfred Ehling / Rolf Porst</i>	
Pretest zur Erhebung „Neukonzeption der laufenden Wirtschaftsrechnungen“	72
 <i>Manfred Ehling / Rolf Porst / Elmar Wein</i>	
Fragebogen-Pretest zur „Erfassung von Produktionstagen und Betriebszeiten“	91
Teilnehmerverzeichnis	116

Günter Kopsch*)

Begrüßung

Ich begrüße Sie herzlich zu der wissenschaftlichen Tagung über „Pretest und Weiterentwicklung von Fragebogen“ hier im Statistischen Bundesamt. Gleichzeitig darf ich Ihnen die Grüße von Herrn Merk, dem Präsidenten des Statistischen Bundesamtes übermitteln, der kurzfristig einen unaufschiebbaren Termin in Bonn wahrnehmen muß.

Die heutige Tagung ist eine gemeinsame Veranstaltung des Arbeitskreises Deutscher Markt- und Sozialforschungsinstitute e. V. (ADM), der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e. V. (ASI) und des Statistischen Bundesamtes. Sie wurde im Rahmen der jährlich stattfindenden Arbeitssitzungen dieser drei Institutionen im Herbst letzten Jahres vereinbart, mit dem Ziel, ein Forum für einen Dialog zwischen Nutzern und Produzenten von statistischen Erhebungen zu schaffen, auf dem über die Möglichkeiten von Qualitätsverbesserungen durch Pretests informiert und diskutiert werden soll.

Als Vertreter der amtlichen Statistik halte ich einen solchen Dialog für außerordentlich wünschenswert. Dies um so mehr, da wir leider nur selten Gelegenheit haben, uns in einem größeren Kreis von Marktforschern, Sozialwissenschaftlern und amtlichen Statistikern gemeinsam über ausgewählte Arbeitsinhalte, die für uns alle gleichermaßen von Interesse sind, auszutauschen. Die Aufgabenstellung, vor der die kommerziellen Marktforschungsinstitute, die sozialwissenschaftlichen Institute und die amtliche Statistik stehen, ist in vieler Hinsicht durchaus verwandt, die Rahmenbedingungen unterscheiden sich jedoch erheblich.

Für uns gilt das Prinzip der Legalisierung, d. h. für jede Bundesstatistik ist – von wenigen Ausnahmen abgesehen – eine Rechtsgrundlage erforderlich. Das erschwert – wie Sie sich denken können – Anpassungen des Programms an neue Anforderungen erheblich. Die sich nun schon seit Jahren hinziehenden Beratungen um die Einführung einer Dienstleistungstatistik sind nur ein Beispiel.

Die kommerziellen Institute müssen sich im Wettbewerb am Markt behaupten, müssen also Gewinne erwirtschaften bzw. Marktanteile sichern und unterliegen somit selbstverständlich auch Restriktionen, nur eben anderen. Die unterschiedlichen Rahmenbedingungen haben auch Auswirkungen auf Ablauf und Organisation der Arbeiten. Doch die Instrumente und statistischen Methoden, derer wir uns alle bedienen, sind im wesentlichen identisch. Es liegt daher nahe, insbesondere in diesen Bereichen zu kooperieren. Die Bundesstatistik ist hierzu bereit.

Unsere heutige Tagung dreht sich um eine für uns alle gleichermaßen grundlegende und zentrale Frage: „Wie können wir die Qualität unserer Erhebungen noch weiter verbessern?“

Heute und morgen geht es weniger um Fragen der Verbesserung durch den Einsatz neuer Techniken und auch nicht um stichprobenmathematische statistisch-methodische Optimierungen, Thema ist vielmehr die Qualitätssteigerung durch Verbesserung der Erhebungsinstrumente mit Hilfe von Pretests, oder genauer: Es geht um Möglichkeiten, systematische

*) Günter Kopsch, Statistisches Bundesamt, Wiesbaden.

Fehler durch Tests der Erhebungsinstrumente im Vorfeld der eigentlichen Feldarbeit – also durch Pretests – zu minimieren.

Der Aufbau der Erhebungsunterlagen, die Formulierung und Reihenfolge der Fragen, aber auch die optische Gestaltung, das Layout des Fragebogens haben erheblichen Einfluß auf die Größe des systematischen Fehlers. Umfang und Art der Antwortausfälle sind in starkem Maße abhängig von der Qualität der Erhebungspapiere. Pretests sind also ein wichtiges Instrument zur Erhöhung der Validität der Daten, sie erhöhen die Sicherheit, auch tatsächlich das zu messen, was gemessen werden soll. Durch diese Validitätserhöhung steigern Pretests bei Erhebungen die Datenqualität und tragen somit dazu bei, Statistik effektiver zu gestalten.

Wir bereiten derzeit eine Testerhebung zur Neukonzeption der Wirtschaftsrechnungen der privaten Haushalte vor. Hierzu wurden gerade in den letzten Wochen die Erhebungsunterlagen einem Pretest unterzogen. Die Ergebnisse dieses vom Zentrum für Umfragen, Methoden und Analysen (ZUMA) in Mannheim durchgeführten Pretests haben uns wichtige Anregungen zur Verbesserung der Fragenbogen geliefert, die inzwischen umgesetzt worden sind.

Betrachtet man zusätzlich zu den Nutzenüberlegungen im Hinblick auf die Datenqualität noch die Kostenseite, so verursachen Pretests ganz offensichtlich zunächst einmal Aufwendungen. Aber Pretests stellen auch ein Rationalisierungs- und Kosteneinsparungspotential dar, denn in der Regel sind nachträgliche Korrekturen von Fehlern in Untersuchungen, bei denen auf Pretests verzichtet wurde – sofern solche Korrekturen nachträglich überhaupt noch möglich sind – kostenintensiver. Die möglichen Einsparungen durch Pretests zeigen sich unter anderem dadurch, daß weniger Nachfragen bei den befragten Haushalten und Unternehmen erforderlich werden. Durch Pretests kann letztlich der Aufwand bei den Plausibilitätskontrollen und der Aufbereitung reduziert und somit die Bearbeitungszeit bis zur Ergebnispräsentation verkürzt werden.

Ein Blick in das Programm unserer heutigen Tagung zeigt, daß es den Organisatoren primär nicht um theoretische Fragestellungen geht, sondern daß ganz konkrete, praktische Erfahrungen und Lösungen bei verschiedensten Pretestverfahren im Mittelpunkt der Veranstaltung stehen. Neben den klassischen Pretesttechniken bilden Fragebogenkonferenzen und Gruppendiskussionen zur Entwicklung, Bewertung und Beurteilung von Fragebogen Themenschwerpunkte. Ebenso finden Ansätze aus dem Bereich der Kognitionspsychologie Eingang und es wird über das Pretestlabor berichtet, das beim Niederländischen Statistischen Zentralamt eingerichtet worden ist. Ich bin überzeugt, daß gerade durch dieses breite Spektrum ein für alle Teilnehmer anregender und fruchtbarer Austausch von Erfahrungen möglich sein wird.

Zum Schluß möchte ich mich schon jetzt bei den Vortragenden für ihre Beiträge und bei Herrn Professor Mohler für die Moderation bedanken. Alle Teilnehmer dieser Veranstaltung werden durch Ihre Beiträge – und die sich anschließenden Diskussionen – Anregungen und Vorschläge erhalten, die wichtige Hilfestellungen bei der Aufgabenerledigung darstellen können.

Pretest und Weiterentwicklung von Fragebogen – Einführung in das Thema

Meine sehr geehrten Damen und Herren,

es ist eine große Ehre, bei dieser wissenschaftlichen Tagung zum Thema „Pretest und Weiterentwicklung von Fragebogen“ die Moderation innezuhaben, und es ist eine große Freude, die einführenden Worte zum Thema sprechen zu dürfen. Es bleiben nur 15 Minuten Zeit dafür, und die wollen genutzt sein. Dennoch gestatten Sie eine kurze Vorrede.

Wir nähern uns einem weiteren Meilenstein auf dem Weg zu einer integrierten, kooperativen Sozialforschung in Deutschland. Wir sind auf diesem Weg schon ein gutes Stück vorangekommen, können zum Beispiel eine in der Welt einzigartige gemeinsame Standarddemographie (Ehling u. a. 1992; Statistisches Bundesamt 1995) vorweisen oder haben mit dem „Datenreport“, zuletzt 1994 erschienen (Statistisches Bundesamt 1994), eine wichtige gemeinsame Datenquelle für die empirische Bearbeitung sozial- und wirtschaftswissenschaftlicher Fragestellungen geschaffen.

Wenn wir uns heute – in bewährter Zusammenarbeit zwischen dem Statistischen Bundesamt, dem Arbeitskreis Deutscher Marktforschungsinstitute e. V. (ADM), der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e. V. (ASI) und dem Zentrum für Umfragen, Methoden und Analysen (ZUMA) – nun Problemen des *Pretests* und der *Fragebogenentwicklung* zuwenden, dann eröffnet dies eine neue Dimension der Kooperation, für die es in anderen Ländern bereits nachahmenswerte Vorbilder gibt.

Wenn wir an dieser Stelle einen Wunsch äußern dürften, wünschten wir uns ein Forum, vergleichbar mit dem der American Association for Public Opinion Research (AAPOR). Die Jahrestagungen der AAPOR sind erfüllt von Diskussionen und Gesprächen über die „Vorgärten“ der universitären, privaten und amtlichen Sozialforschung hinweg. Zum Dinner sitzt man zusammen mit der ehemaligen Direktorin des amerikanischen statistischen Amtes, die jetzt Professorin in Ann Arbor ist, unter deren Leitung hervorragende Sozialwissenschaftler an das Amt berufen wurden und so weiter. Kurz, durch AAPOR ist der Hiatus zwischen akademischer, privater und amtlicher Sozialforschung aufgehoben. Wenn es ein Ziel für die heutige Veranstaltung gibt, dann könnten wir uns ein solches Forum über die Grenzen hinweg als Ziel vorstellen.

Nun aber mitten ins Thema unserer Tagung. Übersicht 1 (siehe S. 8) vermittelt in knapper Form eine Vorstellung davon, womit wir uns im folgenden beschäftigen wollen.¹⁾

*) Prof. Dr. Peter Ph. Mohler, Zentrum für Umfragen, Methoden und Analysen (ZUMA), Mannheim.
Rolf Porst, Zentrum für Umfragen, Methoden und Analysen (ZUMA), Mannheim.

1) Wir danken Peter Prüfer und Margrit Rexroth, beide wissenschaftliche Mitarbeiter in der ZUMA-Feldabteilung, für ihre überaus wertvolle Unterstützung bei der Vorbereitung des Vortrages, der diesem Artikel zugrundeliegt.

Übersicht 1 Inhalt der Einführung in das Thema

Stellenwert und Ziele eines Pretests

Pretestverfahren

- Der „klassische“ Pretest
- Kognitive und Labor-Methoden
- Verfahren ohne Befragungspersonen
- Behaviour Coding

Beginnen wir mit dem Stellenwert und den Zielen eines Pretests.

1 Stellenwert und Ziele eines Pretests

Wo immer man auch hinschauen und nachlesen mag: In der gesamten einschlägigen Literatur werden *Pretests* in allen Projekten der empirischen Sozialforschung als unabdingbare Voraussetzung zur Vorbereitung einer Hauptstudie bezeichnet (Porst 1985, S. 50), zumindest in all jenen Projekten, die zur Datengewinnung Umfragen und Fragebogen einsetzen wollen (mit der Notwendigkeit von Pretests beschäftigt sich der Beitrag von Erwin K. Scheuch in diesem Band, S. 16 ff.).

An dieser Stelle immer wieder gern zitiert der Ratschlag von Sudman und Bradburn (1983, S. 283): „*If you do not have the resources to pilot-test your questionnaire, don't do the study.*“

Dieses Zitat, so eindringlich es wirkt, hat allerdings eine ganz zentrale Schwäche, nämlich: Es setzt Pretesting sehr einschränkend gleich mit Überprüfung eines Fragebogens.

Eigentlich sollte es aber anders sein: Pretests sollten die Aufgabe haben, Hinweise zu liefern über die Funktionsfähigkeit des gesamten Studiendesigns sowie einzelner Bestandteile dieses Designs, und dazu gehört, neben Stichprobenziehung und Stichprobenrealisierung, neben Fragen des Feldes und sogar der Auswertung, eben auch das Befragungsinstrument, der Fragebogen.

In der Praxis ist dies aber nicht der Fall: Zumeist werden alle anderen Aspekte vernachlässigt, und alleine der *Fragebogen* steht im Mittelpunkt des Pretestinteresses. Praktisch handelt es sich beim Pretest also um den Testlauf eines „Fragebogen-Prototyps“, d. h. eines mutmaßlich noch nicht vollständig ausgereiften Fragebogens. Damit wird auch deutlich, in welchen übergeordneten Zusammenhang Pretests einzuordnen sind: Sie sind ein *wesentliches Element im Prozeß der Fragebogenentwicklung*.

Warum wird gerade das Pretesten des *Fragebogens* immer wieder in den Vordergrund gestellt?

Zum einen geschieht dies deshalb, weil der Fragebogen bekanntermaßen eine der zentralen potentiellen Fehlerquellen in Umfragen ist, zum zweiten, weil weitgehender Konsens darüber besteht, daß ein perfekter Fragebogen – wenn es denn einen perfekten Fragebogen überhaupt gibt – nicht am Schreibtisch konstruiert werden kann. Wir stimmen Sudman und Bradburn (1983, S. 283) erneut zu, wenn sie feststellen: „*Even after years of experience, no expert can write a perfect questionnaire.*“

Wir stimmen auch Klaus Allerbeck und Wendy Hoag (1985) zu, die im Zusammenhang mit der Konstruktion der Fragebogen ihrer Jugendstudie sagen: „*Erfahrene Umfrageforscher kennen die hier gültige Variante des Murphyschen Gesetzes, daß alles, was schiefgehen kann, auch schiefgehen wird, schon lange.*“

Nun, da wir alle wissen, wie wichtig ein Pretest ist, suchen wir nach einer Definition, und – oh Wunder – wir finden keine (zumindest keine brauchbare). Trotz des hohen Stellenwerts, der Pretests zugeschrieben wird, gibt es zumindest in der uns bekannten Literatur weder eine allgemein anerkannte Definition noch konkrete und präzise Regeln zur praktischen Durchführung eines Pretests (eine Feststellung, die sich im übrigen mühelos auf andere zentrale Begriffe der empirischen Umfrageforschung übertragen ließe).

Trauen wir uns, eine eigene Definition vorzustellen, so sagen wir:

Pretests sind nichts anderes als die Miniaturausgabe einer beliebigen Form sozialwissenschaftlicher Datenerhebung, wobei sich in der Regel die Konzentration auf die Qualität des Erhebungsinstrumentes richtet.

In Anlehnung an Converse und Presser (1986) sollte ein Pretest im allgemeinen Aufschluß geben über die folgenden Aspekte:

Übersicht 2 Gegenstände eines Pretests

- Probleme des Befragten mit seiner Aufgabe
- Probleme des Interviewers
- Interesse und Aufmerksamkeit des Befragten während des gesamten Interviews
- Wohlbefinden des Befragten (Respondent well-being)
- Interesse und Aufmerksamkeit des Befragten bei einzelnen Fragen
- Verständlichkeit der Fragen
- Reihenfolge der Fragen
- Kontexteffekte
- Häufigkeitsverteilung der Antworten
- Technische Probleme mit Fragebogen/Befragungshilfen
- Zeitdauer der Befragung

Spätestens durch diese Aufzählung wird deutlich, daß es ab jetzt vorrangig um Pretests für mehr oder weniger standardisierte Fragebogen geht.

2 Pretestverfahren

Wenden wir uns der Frage zu, welche unterschiedlichen Pretestverfahren zur Verfügung stehen. Auf der einen Seite haben wir den sogenannten „klassischen“ Pretest, der sich – vereinfacht und verkürzt ausgedrückt – vor allem dadurch kennzeichnet, daß er im Feld durchgeführt wird und der Interviewer die eher passive Funktion eines Beobachters innehat. Auf der anderen Seite haben wir Pretestverfahren, die eher und vorrangig in Labors oder inhouse durchgeführt werden, bei denen sowohl Befragte wie auch Interviewer aktive Rollen beim Pretesten übernehmen (wir fassen sie hier unter kognitive und Labor-Methoden zusammen). Schließlich gibt es Verfahren, die theoretisch oder praktisch ganz ohne Mitwirkung einer Zielperson ablaufen (z. B. Expertenratings, Fragebogenkonferenzen).

Irgendwo zwischen dem klassischen Pretest und dem kognitiven Labor-Pretest ist schließlich die Variante des Behaviour Coding angesiedelt, das bei ZUMA zum Einsatz kommt.²⁾

2.1 Der „klassische“ Pretest

Wenden wir uns zunächst dem klassischen Pretest zu, der auch bezeichnet wird als

- „Conventional pretest“ (Presser, Blair 1994),
- „Standard pretest“ (Oksenberg, Cannell, Kalton 1991),
- Beobachtungspretest (ZUMA), oder, schon eher etwas respektlos, als
- „Old-style-pretest“ (Fowler 1992).

Die Bezeichnungen „klassisch“, „konventionell“ und „old-style“ deuten darauf hin, daß diese Attribute dem traditionellen Beobachtungspretest erst in jüngster Zeit – genauer, seit der Entwicklung alternativer Verfahren – verliehen worden sind.

Auch für die Durchführung eines „klassischen Pretests“ existieren keine verbindlichen bzw. allgemein akzeptierten Regeln. Zwar findet sich in jedem sozialwissenschaftlichen Methodenlehrbuch ein Kapitel zum Thema „Pretest“, im Vergleich miteinander werden jedoch höchst unterschiedliche und zum Teil auch widersprüchliche Empfehlungen bezüglich der Durchführung eines Pretests gegeben. Blättert man die Methodenliteratur durch, so stößt man auf Uneinigkeit zum Beispiel bei

- der notwendigen Fallzahl (variiert zwischen $N = 10$ und $N = 200$),
- dem Einsatz der Interviewer (erfahrene, speziell ausgebildete Pretest-Interviewer oder Querschnitt aller Interviewer),
- der Informiertheit der Befragten („Participating pretest“ = Befragter ist über Testcharakter informiert; im Gegensatz dazu „Undeclared pretest“ = Befragter ist nicht informiert, agiert unter realistischen Bedingungen),
- der Terminologie an sich (Pretest, Vortest, Testbefragung, Pilot study, Question testing, Trial run etc.).

2) Für die Weiterentwicklung des modifizierten Behaviour Coding sowie generell die Weiterentwicklung von Pretestverfahren sind bei ZUMA vorrangig Margrit Rexroth und Peter Prüfer zuständig. Eine umfassende Veröffentlichung von ZUMA zum Thema Pretesting liegt vor (Prüfer, Rexroth 1996).

Obwohl es nirgendwo explizit geschrieben steht, gibt es doch zumindest eine Art Übereinstimmung darüber, wie das „Grundgerüst“ eines „klassischen“ Pretests beschaffen ist bzw. sein sollte:

Übersicht 3

Grundgerüst eines „klassischen“ Pretests

- Einmaliges Testen des Fragebogens unter möglichst realistischen Hauptstudien-Bedingungen
- Durchführung von 20 bis 70 Interviews (Quota oder Random)
- Interviewer haben die Aufgabe, Probleme und Auffälligkeiten bei der Durchführung der Interviews zu beobachten und zu berichten.
- In der Regel passives Verfahren, d. h., der Interviewer beobachtet nur („Beobachtungspretest“), ohne aktiv zu hinterfragen.
- Zugrundeliegendes Prinzip: Man versucht, aus der Reaktion/Antwort des Befragten Rückschlüsse auf sein Fragenverständnis zu ziehen.

2.2 Kognitive und Labor-Methoden

Im Laufe der achtziger Jahre wurde von Kognitionspsychologen in den USA eine Reihe von Techniken zur Untersuchung des Frage-Antwort-Prozesses entweder neu entwickelt, oder bereits vorhandene Techniken wurden zu neuem Leben erweckt. Dabei ging es also zunächst einmal gar nicht um Pretests. Die Eignung dieser Techniken zu Pretest-Zwecken wurde im Grunde genommen erst später erkannt (Kognitionspsychologische Ansätze sind Gegenstand des Beitrags von Michaela Wänke in diesem Band, S. 46 ff.; über Labormethoden lesen wir bei Hans Akkerboom, S. 66 ff.; von daher wollen wir auch hier nur einen ersten Überblick geben und auf eine Wertung verzichten).

Zu den kognitiven Techniken im einzelnen zählen wir:

Übersicht 4 **Kognitive und Labormethoden**

Think-Aloud-Technik mit zwei Varianten:

- Concurrent Think Aloud
- Retrospective Think Aloud

Probing-Technik mit den wichtigsten Varianten:

- Follow-Up-Probing
- Post-Interview-Probing
- Comprehension Probing
- Information Retrieval Probing

Memory Cues**Confidence Ratings****Paraphrasing****Sorting**

- Free Sort
- Dimensional Sort

Think Aloud ist eine Neuentwicklung, bei der die Befragungsperson aufgefordert wird, „laut zu denken“. Die Think-Aloud-Technik entstammt der stark kognitionspsychologisch orientierten Gedächtnisforschung und ist vor allem bei retrospektiven Fakt-Fragen sinnvoll einsetzbar. Es gibt zwei wichtige Varianten: Beim *Concurrent Think Aloud* denkt der Befragte laut, während er seine Antwort formuliert (z. B. Cannell u. a. 1981; Jabine u. a. 1984; Loftus 1984), beim *Retrospective Think Aloud* denkt der Befragte nach der Beantwortung der Frage laut darüber nach, wie seine Antwort zustande gekommen ist.

Eine eher alte Technik ist das *Probing*. Unter Probing verstehen wir das explizite Hinterfragen der Antwortstrategien der Befragungsperson. Die wichtigsten Varianten sind:

- | | |
|---|--|
| – <i>Follow-Up-Probing</i> : | Nachfrage sofort nach der Antwort; |
| – <i>Post-Interview-Probing</i> : | Nachfragen nach Beendigung des Interviews; |
| – <i>Comprehension Probing</i> : | Nachfrage zum Fragenverständnis; |
| -- <i>Information Retrieval Probing</i> : | Nachfrage zur Informationsbeschaffung. |

Vier weitere Neuentwicklungen im Bereich kognitiver und Labor-Methoden sind:

- *Memcry Cues*: Erinnerungshilfen für den Befragten
- *Confidence Ratings*: Befragter bewertet den Grad der Verlässlichkeit seiner Antwort.
- *Paraphrasing*: Befragter soll die Frage mit eigenen Worten formulieren.
- *Free/Dimensional Sort*: Befragter gruppiert vorgegebene Items entweder nach eigenen (Free Sort) oder vorgebenden Kriterien (Dimensional Sort) anhand vorgegebener Skalen.

Schließlich haben wir hier noch das *Focus Interview* bzw. die *Focus Group*, eine im Grunde alte Technik, nämlich die relativ unstrukturierte Diskussion über das Befragungsinstrument mit Gruppen von Befragungspersonen oder mit einzelnen Befragten. Die Focus Group ist sozusagen das methodische Äquivalent zur Gruppendiskussion (vgl. den Beitrag von Klaus Haupt in diesem Band, S. 35 ff.).

Kennzeichnend für alle diese Techniken ist, daß sie einerseits Test-Charakter besitzen, andererseits aber auch gleichzeitig Instrumente zur Fragebogenentwicklung darstellen. Weiterhin ist festzuhalten, daß sich diese Techniken vorwiegend mit dem Test einzelner oder weniger Fragen beschäftigen und nicht mit dem Fragebogen als Ganzem.

Dies bedeutet, daß wir Labor-Techniken zwar zur Vorbereitung einer Befragung empfehlen und einsetzen, daß sie aber unserer Ansicht nach nicht den Test des gesamten Instruments – in welcher Form auch immer – ersetzen können.

Zumindest einige dieser „Labor“-Methoden sind auch im „normalen“ Interview einsetzbar, z. B. wurde die Think-Aloud-Methode von ZUMA bereits in Pretest-Interviews im Feld erfolgreich angewandt (Prüfer, Rexroth 1996).

2.3 Verfahren ohne Befragungspersonen

Bei den Verfahren, die im Prinzip ohne Befragungsperson auskommen, ist vor allem das *Expertenrating* zu nennen: Mehrere Experten stellen zunächst unabhängig voneinander die ihrer Ansicht nach vorhandenen Problemé eines Fragebogens fest und diskutieren anschließend gemeinsam darüber. Eine prominente Form des Expertenratings finden wir in der Fragebogenkonferenz des Instituts für Demoskopie (vgl. den Beitrag von Elisabeth Noelle-Neumann in diesem Band, S. 55 ff.).

2.4 Behaviour Coding

Bleibt schließlich das Behaviour Coding. Das grundlegende Prinzip dieses Verfahrens ist die *Klassifizierung von Verhalten*.

Behaviour Coding wurde ursprünglich eingesetzt, um Interviewerverhalten zu bewerten (z. B. Cannell, Lawson, Hausser 1975; Kreiselmaier u. a. 1989); in späteren Arbeiten wurde es dann auch auf das Befragtenverhalten angewandt (z. B. Morton-Williams 1983; Cannell, Kalton, Fowler 1985; Prüfer, Rexroth 1985).

Beim traditionellen Behaviour Coding bewerten *Coder* nach dem Interview das Befragtenverhalten mittels eines Codesystems, das mehr oder weniger umfangreich sein, d. h., mehr oder weniger differenziert Verhalten erfassen kann. Der quantitative Häufigkeitswert

adäquater und inadäquater Befragtenverhaltensweisen bei einer Frage wird als Qualitätsindikator dieser Frage gewertet, weil man davon ausgeht, daß bei inadäquatem Befragtenverhalten die Frage irgendeine Schwierigkeit, irgendein Problem aufweist.

In der ZUMA-Feldabteilung wird das Verfahren seit Jahren modifiziert eingesetzt (Prüfer, Rexroth 1985). Die Modifikation besteht in einer Verbindung zwischen traditionellem Behaviour Coding und Beobachtungspretest. Wie sieht diese Verbindung aus?

Die Bewertung wird nicht vom Coder nach dem Interview, sondern vom *Interviewer während des Interviews* vorgenommen. Bedingung ist, daß das Codesystem auf das Notwendigste reduziert ist. Das heißt, das Befragtenverhalten wird mittels einer Codeziffer nur im Hinblick darauf bewertet, ob es im Sinne der Fragestellung adäquat oder nicht adäquat ist. *Zusätzlich* beschreibt der Interviewer für jede inadäquate Verhaltensweise in einem Interviewer-Erfahrungsbericht nach dem Interview möglichst detailliert das Befragtenverhalten.

Der *Vorteil dieses Vorgehens* besteht darin, daß man neben dem quantitativen Häufigkeitswert als Qualitätsindikator auch die konkrete Ursache für den Qualitätsmangel einer Frage kennt. Also: Quantifizierung von Verhaltensweisen plus qualitative Beschreibung des Problems (Prüfer, Rexroth 1985).

In der Literatur wird auf diese Art der Anwendung des Behaviour Codings nicht verwiesen und unseres Wissens wird es außer bei ZUMA in der Praxis so auch nicht angewendet. Für den Interviewer bedeutet das modifizierte Behavior Coding eine hohe Anforderung, der er nur aufgrund problemorientierter intensiver Schulungsmaßnahmen gerecht werden kann. Das Verfahren hat sich bei ZUMA als sinnvolle Variante des Pretesting bewährt.

Literaturhinweise

Allerbeck, K., Hoag, W. (1985): Zur Methodik der Umfragen, Frankfurt am Main.

Cannell, Ch. F., Kalton, G. W., Fowler, F. W. (1985): Techniques for Diagnosing Cognitive and Affective Problems in Survey Questions, Michigan.

Cannell, Ch. F., Lawson, S. A., Hausser, D. L. (1975): A Technique for Evaluating Interviewer Performance, Michigan.

Cannell, Ch. F., Miller, P. V., Oksenberg, L. (1981): Research on Interviewing Techniques, Sociological Methodology, San Francisco.

Converse, J. M., Presser, S. (1986): Survey Questions. Handcrafting the Standardized Questionnaire, Sage University Paper Series on Quantitative Applications in the Social Sciences, Beverly Hills.

Ehling, M., von der Heyde, Ch., Hoffmeyer-Zlotnik, J., Quitt, H. (1992): Eine deutsche Standarddemographie, in: ZUMA-Nachrichten, 31, Mannheim, S. 29 – 46.

Fowler, F. J. (1992): How Unclear Terms Affect Survey Data, in: Public Opinion Quarterly, 56, S. 218 – 231.

Jabine, Th. B., Straf, M. L., Tanur, J. M., Tourangeau, R. (Hrsg., 1984): Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines, Washington D.C.

Kreiselmaier, J., Prüfer, P., Rexroth, M. (1989): Der Interviewer im Pretest. Evaluation der Interviewerleistung und Entwurf eines neuen Pretestkonzepts, ZUMA-Arbeitsbericht Nr. 89/14, Mannheim.

Loftus, E. (1984): Protocol Analysis of Responses to Survey Recall Questions, in: Th. B. Jabine, M. L. Straf, J. M. Tañur, R. Tourangeau (Hrsg.): Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines, Washington D.C., S. 61 – 64.

Morton-Williams, J. (1983): The Use of „Verbal Interaction Coding“ for Evaluating a Questionnaire, in: Quality and Quantity, 13, S. 59 – 75.

Oksenberg, L., Cannell, Ch. F., Kalton, G. (1991): New Strategies for Pretesting Survey Questions, in: Journal of Official Statistics, 7, S. 349 – 365.

Porst, R. (1985): Praxis der Umfrageforschung. Erhebung und Auswertung sozialwissenschaftlicher Umfragedaten, Teubner-Studienskripten zur Soziologie 126, Stuttgart.

Presser, S., Blair, J. (1994): Do Different Methods Produce Different Results? Sociological Methodology, 23.

Prüfer, P., Rexroth, M. (1985): Zur Anwendung der Interaction-Coding-Technik, in: ZUMA-Nachrichten, 17, Mannheim, S. 2 – 49.

Prüfer, P., Rexroth, M. (1996): Pretesting: Überblick und Trends, ZUMA-Arbeitsbericht, Mannheim.

Statistisches Bundesamt (Hrsg., 1994): Datenreport 1994. Zahlen und Fakten über die Bundesrepublik Deutschland. In Zusammenarbeit mit dem Wissenschaftszentrum Berlin für Sozialforschung und dem Zentrum für Umfragen, Methoden und Analysen, Mannheim, Schriftenreihe Band 325, Bundeszentrale für politische Bildung, Bonn.

Statistisches Bundesamt (Hrsg., 1995): Demographische Standards. Eine gemeinsame Empfehlung des Arbeitskreises Deutscher Marktforschungsinstitute (ADM), der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute (ASI) und des Statistischen Bundesamtes, Reihe Methoden – Verfahren – Entwicklungen, Materialien und Berichte, 2. überarbeitete und erweiterte Auflage, Wiesbaden.

Sudman, S., Bradburn, N. M. (1983): Asking Questions. A Practical Guide to Questionnaire Design, 2. Auflage, San Francisco.

Die Notwendigkeit von Pretests zur Vorbereitung statistischer Erhebungen

„If you do not have the resources to pilot-test your questionnaire, don't do the study“ – so Seymour Sudman und Norman Bradburn (1983). Eine Selbstverständlichkeit! Oder nicht?

1 Pretests auch bei vertrauten Erhebungen

In einem bedeutenden Rundfunksender kam es im März 1995 zu einer Kontroverse. Der Zeitschrift für die Zuschauer des TV-Programms sollte ein Fragebogen beigelegt werden. Das mit der Intendanz abgestimmte Fragenprogramm war mit neun Seiten zu umfangreich geraten und mußte auf sechs Seiten zusammengestrichen werden. Damit war auch eine Modifikation von Fragen verbunden. Als nun wissenschaftliche Mitarbeiter der Medienforschung dieses Senders die Modalitäten eines Pretests verabreden wollten, widersprach ihnen eine seit vielen Jahren in dieser Abteilung tätige Sachbearbeiterin auf das heftigste. Das Kürzen des Fragebogens als Gemeinschaftsarbeit sei doch bereits der Pretest gewesen. Man kenne doch ausreichend sein Publikum.

Gerade bei wiederkehrenden Erhebungen mag sich ein solches Gefühl ausbreiten, hier gebe es nichts neu zu lernen. Statistische Erhebungen sind von dieser Art – ja bei manchen Erhebungen (Volkszählungen, Mikrozensus) wird das Fragenprogramm sogar gesetzlich fixiert. Aber: Wenn ohne Pretest geforscht wird, gleicht dies einem Blindflug, selbst wenn es der Forschung um eine konventionelle Erhebung geht.

Auch bei gewohnten Bedingungen sind Pretests unerlässlich. Selbst wenn sich die einzelnen Fragen über längere Zeit hinweg bewährt haben, so ist ihr Zusammenfügen in einen Fragebogen ein überprüfungsbedürftiger Vorgang¹⁾. Es geht ja keinesfalls nur um die Beurteilung des Fragebogens für sich, sondern auch um die Anwender und die Adressaten des Instrumentes. Entsprechend schreibt Galtung bereits 1967: *„It should be noticed that the Pretest is never in any abstract sense only a test of the instrument but a test of the entire process of data collection and even of the first steps of analysis. In general what we do in a pretest analysis is to check whether the bridge between the problem and reality has been constructed.“* (Galtung 1967, S. 137–138.)

Eine Gesellschaft wie die unsrige ist sowohl heterogen wie auch in fortwährender Bewegung. Die Heterogenität hat zur Folge, daß niemand von uns über alle Soziallagen und Untergruppen der Bevölkerung hinweg ausreichend kundig ist. Und die Veränderungen haben zur Folge, daß die gleiche Frage morgen anders wirken kann als heute. Noch größer ist die Unsicherheit bei internationalen Erhebungen.

Vor einem Eingehen auf einzelne Aspekte ist an eine Grundeinsicht für Erhebungen bei menschlichen Populationen zu erinnern: Die Konstruktion eines angemessenen Erhebungsbogens ist zwar wissenschaftlich begründet, aber doch zugleich Kunsthandwerk – im Engli-

^{*)} Prof. Dr. Erwin K. Scheuch, Universität zu Köln, Kölner Gesellschaft für Sozialforschung.

¹⁾ „In jedem Fall (!) sollte ein auch nur in Teilen neu konstruiertes Erhebungsinstrument einem Pretest unterzogen werden“ – so Diekmann (1995), S. 169.

schen: „it is a craft“ – und zwar ein sehr anspruchsvolles. Bei allen komplizierteren Verfahrensweisen im Testen von Bogen darf das nicht vergessen werden. Ein Handwerk und dessen Kunstfertigkeiten sind immer wieder durch praktische Bewährung weiterzuentwickeln.

Noch einmal seien Sudman und Bradburn zitiert: *„Even after years of experience, no expert can write a perfect questionnaire. Between us we have more than fifty years of experience in questionnaire construction, and we have never written a perfect questionnaire on the first draft, nor do we know any professional social scientists who claim that they can write questionnaires that need no revision.“*²⁾

Im Jahre 1953 führten wir selbst repräsentative Befragungen in Köln mit dem Thema „Ein Interview über das Interview“ durch. Nebeneinander wurden drei verschiedene Stichproben hinsichtlich der Praktikabilität einer Umsetzung von Stichprobenplänen überprüft: Random Walk, Adressensample und Quotenverfahren. Neben der Frage nach der Praktikabilität von Auswahlplänen wurde auch versucht zu ermitteln, aus welcher Perspektive heraus welche Fragen vom Interviewten beantwortet wurden. Dabei zeigte sich zwar eine große Heterogenität, aber auch zugleich ein Leitmotiv: Das Interview wurde als Gelegenheit verstanden, „denen da oben“ die Sicht von „uns da unten“ mitzuteilen. Es erwies sich damit als sehr fördernd für die Motivation zur Mitarbeit, einige Fragen einzubauen, die für den Forscher von geringer Bedeutung waren, aber für den Befragten Gelegenheit gaben, seinen Ärger mit dem Lauf der öffentlichen Angelegenheiten auszudrücken.

Dem Programm der Tagung ist zu entnehmen, daß es eine ganze Anzahl von wichtigen Beiträgen zu sehr speziellen Problemen des Pretests geben wird. Angesichts dessen erschien es sinnvoll, an dieser Stelle zunächst eine Übersicht zu versuchen und dabei insbesondere den Bezug einzelner Problemstellungen zur Forschung als Prozeß herzustellen.

2 Zur Problematik der Sprache

Bei der Umsetzung eines Fragenprogramms stellt sich als erstes Problem die Formulierung der Fragen. Analytisch ist hier die gleiche Aufgabe zu lösen wie in der Massenkommunikation. Eine Gesellschaft wie die deutsche ist heterogen, und in ihr sind nebeneinander für verschiedene Situationen und unterschiedliche Personenkreise unterschiedliche Arten der Verwendung der gleichen deutschen Sprache üblich. Daneben gilt auch für uns, daß im Alltag die Sprache mehrdeutig ist, weil sie eben multifunktionalen Zwecken dient. Sie dient vor allem nicht nur dem Austausch von Kognitionen, sondern weithin der Übermittlung von Wertungen. Hinzu kommen die Ambivalenzen, die wir gegenüber vielen Sachverhalten empfinden, wofür die Schwierigkeiten mit Worten wie Gesetz oder Beamte ein besonderes Beispiel sind. Dennoch sollen Fragen möglichst einheitlich und im Sinne des Mitteilenden verstanden werden.

Bei Untersuchungsgegenständen wie in der Politik ist das in den meisten Fällen nicht besonders schwierig, weil für den politischen Diskurs die Medien die Spracharten und -ebenen ziemlich standardisiert haben. Anders ist es etwa bei dem Thema Sexualität, wo es

2) Norman Bradburn ist immerhin der Leiter des „National Opinion Research Center“. NORC ist das einzige nicht-kommerzielle Umfrage-Zentrum, das regelmäßig Repräsentativbefragungen in den gesamten USA durchführt. Es gilt international als Spitzeninstitut.

oft sehr schwierig ist, zwischen den vulgären Bezeichnungen des privaten Lebens und medizinischen Termini eine neutral-alltagssprachliche Ausdrucksform zu finden. Befragt man etwa Arbeiter über ihr privates Verhältnis zu Kollegen, so muß man in Deutschland im Interview die Hochsprache für eine Sache verwenden, die vom Befragten selbst in den meisten Teilen des Landes im Dialekt erlebt wird. Mit der Verwendung der Hochsprache ist dann offensichtlich eine Verfremdung des sprachlich Ausgedrückten im Vergleich zum unmittelbaren Erlebnis verbunden. Die Problematik ist allerdings im Laufe der Zeit geringer geworden, indem sich eine Art „Illu-Deutsch“ – also eine Illustriertensprache – für die verschiedensten Lebensbereiche durchgesetzt hat.

Wo keine unmittelbare Anschauung vorliegt, keine berufliche oder anderweitige Erfahrung mit dem Gegenstand der Befragung, bleibt Verständlichkeit der Sprache ein großes Problem. Wir wissen aus zahlreichen Pretests, daß weit verbreitet unter solchen Begriffen wie „Dividenden“ oder sogar „Nettoeinkommen“ Dinge verstanden werden, die mit der wirklichen Bedeutung des Terminus nur begrenzt etwas zu tun haben. Wirtschaftliche Vorgänge werden nun einmal von der ganz großen Mehrheit der Bevölkerung nur schemenhaft verstanden.

Bei aller Betonung der Problematik von Sprachebene und situationsgebundenem Verständnis befinden wir uns jedoch in den westeuropäischen Ländern in einer verhältnismäßig günstigen Situation zunehmender Vereinheitlichung von Sprachversionen. Die Unterschiedlichkeit der Sprachen innerhalb einer Kultur ist in sehr vielen traditionellen Gesellschaften bzw. in Gesellschaften mit starken vormodernen Überlieferungen unvergleichlich größer. Da gibt es eben verschiedene als korrekt zu berücksichtigende Sprachen, je nachdem, ob gesellschaftlich von oben nach unten geredet wird, von alt zu jung oder von Frau zu Mann. Aus Java wird berichtet, daß es 16 verschiedene Bezeichnungen für die Ehefrau des Gesprächspartners gibt. Die Anwendung der jeweiligen Bezeichnung ist abhängig davon, ob der Fragende jünger oder älter als sein Gegenüber ist, nur bekannt oder befreundet, höheres oder geringeres Prestige besitzt etc. (Scheuch 1973, S. 81).

3 Die Einzelfrage als Gegenstand der Überlegungen

Häufig wird in Umfragen die folgende Frage gestellt: „Ist zur Zeit in Ihrer Familie jemand krank?“ So gestellt ist diese Frage keineswegs eindeutig. Je nach Zusammenhang wird vom gleichen Befragten die Familie unterschiedlich abgegrenzt. Für Sachverhalte, die sich im Kontext der alltäglichen Geselligkeit ereignen, ist diese Abgrenzung anders, als wenn die Möglichkeit der Hilfeleistung Thema ist. Wir wissen aus zahlreichen Netzwerkuntersuchungen, daß im letzteren Fall Familie als nähere Verwandtschaft verstanden wird und nicht nur, wie bei dem ersten Sachverhalt, als Wohngemeinschaft. Es versteht sich übrigens zusätzlich von selbst, daß auch der Begriff „krank“ insbesondere nach Alter unterschiedlich abgegrenzt wird. Für Zwecke des Pretests eignen sich in solchen Fällen Nachfragen über das, was der Befragte mit seinen Antworten an Vorstellungen verbunden hat. Allerdings kann solches Nachfragen nicht fortwährend erfolgen, weil damit die Unbefangenheit des Interviews gestört ist.

Öfters handelt es sich jedoch nicht nur um die Verständlichkeit von Worten, sondern darüber hinaus um eine sachliche Überforderung des Befragten als Beobachter seiner selbst und der Umwelt. Ein Beispiel sind Zeitbudgetuntersuchungen, die bei mündlicher Befragung nur

grobe Zeitblöcke einigermaßen verlässlich vermitteln können. In Fragebögen für persönliche Interviews gilt es dann, die zeitlichen Erlebnisräume als Einheiten der Selbstbeobachtung zu treffen – etwa abgegrenzt jeweils durch Ortsveränderung bzw. Wechsel der Partner. Das reicht vielen Zeitbudgetforschern nicht, die lieber solch kurze Beobachtungsräume wie jeweils 5 Minuten ermittelt hätten. Hier wird durch Pretest zu prüfen sein, ob nicht zusätzlich zur mündlichen Befragung ein anschließendes Selbsteintragen in einen zurückgelassenen Beobachtungsbogen angezeigt ist.

Das führt hin zu einer allgemeineren Überlegung: Welche Art von Evidenz soll durch die Antwort ausgedrückt werden? Meist ist im Interview der Befragte als Beobachter seiner selbst angesprochen, daneben selbstverständlich auch als Beobachter seiner Zeitgenossen – woraus sich dann das Problem ergibt, unter welchen Umständen er hierüber Beobachtungen überhaupt sammeln konnte. Nicht selten ist aber die Frage ein bloßer Stimulus, der nicht der in der Frage scheinbar angesprochenen Information über einen Sachverhalt dient, sondern eine Bewertung ermitteln will. Beispiele: „Wie hoch schätzen Sie die Zahl derjenigen ein, die durch Verkehrsunfälle in einem Jahr auf unseren Straßen sterben? Und wie häufig sind Ihrer Meinung nach dazu im Vergleich tödliche Unfälle im Haushalt?“ Sachlich besteht hier kein Informationsbedarf, da wir bereits wissen, daß der Haushalt gefährlicher ist als der Straßenverkehr. Indem aber der Befragte selbst Mutmaßungen mitteilt, erhalten wir Urteile über die relative Verbreitung von Besorgnissen. Dies ist eine durchaus wichtige Information, da es nur einen schwachen Zusammenhang zwischen Quellen für Besorgnis und tatsächlicher Gefährdung gibt.

Während üblicherweise im Pretest geprüft wird, ob die Fragen auch neutral in dem Sinn sind, daß sie keine Deutung nach der einen oder anderen Richtung hin nahelegen, kann der Forschungszweck auch eine bewußte „Ladung“ von Fragen nahelegen – etwa um die Stabilität einer Einstellung zu prüfen. Hier gibt es eine „klassische“ Erfahrung aus dem Jahre 1941. Die neutrale Fassung einer Frage lautete: „Sollten die Vereinigten Staaten jetzt in den Krieg eintreten?“ Wurde der Frage die Wendung „... und eine Armee nach Europa entsenden“ hinzugefügt, sank der Anteil der Befürworter von 17 auf nur 7 % (Cantril 1951). Die beiden Formulierungen bedeuteten zwar der Sache nach das gleiche; denn was für einen Sinn hatte der Kriegseintritt ohne die Entsendung von Soldaten gegen den Gegner? Die emotionale Wirkung war aber verschieden, weil dem Antwortenden die Konsequenz einer Bejahung genannt wurde, die ihm ohne die Hinzufügung nicht so präsent gewesen wäre. Daraus schlossen die Sozialforscher, daß die Befürwortung eines Kriegseintritts selbst bei der Minderheit der Zustimmenden keine sehr belastbare Einstellung sei.

4 Das Pretesten des Fragebogens

Der entscheidende Abschnitt in der Entwicklung eines Erhebungsinstruments in der Umfrageforschung ist der Test des gesamten Fragebogens. Verschiedene Zwecke werden hiermit verfolgt:

- Welchen Zeitaufwand erfordert eine einzelne Frage relativ zu den anderen Fragen?
- Werden die Fragen akzeptiert?
- Werden alle Fragen verstanden?
- Funktioniert die Filterführung?

- Gibt es Kontexteffekte, wodurch eine vorausgegangene Frage auf weitere wirkt?
- Welche Fragen sind ergiebig für den Forschungszweck?

Neben der Frageformulierung sind dann insbesondere auch die Antwortvorgaben zu prüfen bzw. – wenn die Frage offen gestellt wird – empirische Unterlagen für die Rechtfertigung einer Reduzierung von Antwortmöglichkeiten auf zu entwickelnde Vorgaben heranzuziehen. Keine dieser hier aufgeführten Zielsetzungen kann verlässlich nur aufgrund der eigenen bisherigen Erfahrung als Forscher beantwortet werden. So ist es üblich, daß der Zeitaufwand einzelner Fragen im voraus falsch beurteilt wird. Immer wieder wird die Erfahrung gemacht, daß wir als Forscher mehr wissen wollen, als die verfügbare Zeit zuläßt. Bekanntlich sinkt ja mit zunehmendem Zeitaufwand eines Interviews die Aufmerksamkeit des Befragten für seine Antwort. Jedenfalls gehört das Kürzen von Fragebogen aufgrund der im Feld gemachten Erfahrungen zum Alltag der Forschung.

Auf einige Techniken zur Prüfung, ob eine Frage verstanden wird, soll anschließend noch verwiesen werden. Einfacher ist es für den Interviewer zu entscheiden, ob die Frage allgemein auf Akzeptanz trifft. Umfrageforscher haben einen ganzen Katalog an Tricks entwickelt, um die Akzeptanz sogenannter „schwieriger Fragen“ zu verbessern (sie werden oftmals ironisiert mit dem Beispiel: „In sehr vielen Familien läßt es sich nicht vermeiden, öfters die Frau zu verprügeln. Wie ist das bei Ihnen: Wie oft verprügeln Sie Ihre Frau?“). Wie dieses persiflierende Beispiel offenkundig werden läßt, haben solche Tricks ihre Grenzen, die erst empirisch zu ermitteln sind.

In heterogenen Bevölkerungen wie der deutschen ist ein Fragenprogramm immer nur für einen Teil von Bevölkerungen relevant. Das Führen des Interviews um Untergruppen herum, die bei dem betreffenden Satz an Fragen nicht berücksichtigt werden sollen, ist besonders anfällig für Fehler sowohl beim Design wie auch bei der späteren Handhabung. Selbst wenn der Entwerfer eines Fragebogens sicher ist, bei der Filterführung keine Fehler begangen zu haben, bedarf es doch einer empirischen Prüfung, ob das beim Interviewer auch wirkt wie erwartet. Institute pflegen auf das Layout ihres Fragebogens stolz zu sein und dieses Layout als ein Markenzeichen ihres Instituts zu sehen. So ansprechend das Layout auch sein mag, so ist es doch belanglos, wenn der Interviewer damit nicht umgehen kann.

Als Halo-Effekt (übersetzt: Ausstrahlungswirkung) versteht man, wenn einzelne Fragen oder Fragegruppen die Beantwortung der folgenden Fragen beeinflussen. Solche Effekte sind immer wieder nachgewiesen worden, und gelegentlich gibt es keinen Weg, sie durch Formulierungskünste und Plazierungsartistik völlig auszuschließen. Dann hilft nur noch ein Split-Ballot, um nachträglich die Größe der Verzerrung anzugeben – selbst wenn man dann gelegentlich immer noch nicht die Richtung weiß.

Einzelfragen sind nicht die Regel, sondern Fragegruppen. Als Abfolge wird häufig das „Trichtern“ (funnelling) gewählt: Von einer allgemeinen Formulierung ausgehend werden dabei die Fragen immer spezieller (Kromrey 1983, S. 206). Auch die Umkehrung dieser Ordnung ist nicht selten, wenn eine Kenntnis des Spezifikums unterstellt werden kann und zu einer Verallgemeinerung hingeführt werden soll (revers funnelling). Hier ist ganz besonders prüfbedürftig, ab welchem Befragungsschritt die Antwortenden nicht mehr folgen.

Ein für uns beeindruckendes Beispiel dafür, daß ohne Testen der Ertrag einer Fragegruppe nicht eingeschätzt werden kann, erbrachte unlängst eine Untersuchung der Wirtschaftselite. Eines der Untersuchungsthemen waren Karriereverläufe, und wir verwandten eine ganze Anzahl von zeitraubenden Fragen, um dann später diese Karriere formal als Markow-Kette abbilden zu können. Dieses komplizierte Unterfangen erwies sich als unnötig, weil die Karriereverläufe der Spitzenkräfte durchweg von großer Einfachheit waren. Wir konnten uns schätzungsweise zehn Minuten Befragungszeit sparen.

Aus psychologischen Tests ist ein Mechanismus belegt, der auch bei Umfragen zu prüfen ist. Mit „Response set“ ist gemeint, daß Fragefolgen in einer stereotypen Weise beantwortet werden, und zwar unabhängig von dem Inhalt der jeweils gestellten Fragen. Aus Frankreich wird berichtet, daß es hier eine Tendenz zur Verneinung gebe, wie andererseits für Amerika eine Neigung zum Bejahen behauptet wird. Stehen mehrere Antwortkategorien für den Befragten zur Auswahl, so besteht in Deutschland eine Tendenz, mittlere Kategorien zu bevorzugen (Kriz 1981). Nicht zuletzt muß der Interviewer berichten, ob Ermüdungserscheinungen eintreten. Gerade bei Marktforschungen mit sehr monotonen Fragestellungen und auch bei amtlichen Erhebungen, deren Zweck den Befragten wenig interessieren, sind Ermüdungen mit ihrer Minderung der Antwortqualität häufig. Dann empfiehlt es sich, in den Bogen „interessante Fragen“ einzubauen.

Umgekehrt sind unsere eigenen persönlichen Erfahrungen. Für Universitätserhebungen durch private Institute ist es nützlich, einige demographische und administrative Fragen an den Anfang zu stellen, um damit die auch positive Anmutung einer (beinahe) amtlichen Erhebung zu verstärken.

Schließlich ist noch ein Sonderproblem beim Pretesten zu bedenken: die Notwendigkeit von Übergangsfragen und deren Wirkung. Insoweit in einem Fragebogen eine ganze Reihe von Themen angesprochen werden – und das ist die Regel –, müssen die Übergänge für den Befragten gleitend werden. Sonst wirkt das Springen von Thema zu Thema von irritierend bis hin zu desorientierend.

5 Die Lehre vom Pretesten

Eine Lehre vom Pretesten analog zu dem Empfehlungskatalog für das Formulieren guter Fragebogen gibt es nicht. Es gibt Usancen sowie einige besondere Techniken, die interessant sind, aber keine allgemeine Anwendung finden. Hier sind viele Fragen offen, wie auch eine ganze Reihe von Punkten strittig geblieben sind. Mit der wichtigste ist: Soll der Interviewer den Sinn der einzelnen Frage kennen oder nur den Informationsstand erhalten, über den später auch die Interviewer der Haupterhebung verfügen? Wir selbst schließen uns hier Marie Jahoda und ihren Mitarbeitern an: *„Those who do the interviewing on the pretest must have a clear understanding of the purpose of the study.“* (Marie Jahoda 1981.) Weithin unstrittig ist dagegen, daß das Pretesten den erfahrensten und besten Interviewern übertragen werden sollte.

Das eindeutig vorherrschende Verfahren beim Pretesten ist das interpretierende Nachfragen (interpretative probing). Wann immer Interviewer beobachten, daß der Antwortende Probleme mit einer Frage hat, muß er nachfassen. Beispiele sind: „Wie haben Sie das gemeint?“ Oder: „Sind Sie da sicher?“ Oder: „Können Sie darüber noch etwas mehr sagen?“ Das ist

selbstverständlich nicht im voraus standardisierbar und ist damit als qualitatives Vorgehen zu werten. Die Interviewer sind entsprechend anzuhalten, ein Protokoll zu führen, aus dem möglichst gut das Problem mit den betreffenden Fragen zu erschließen ist. Gute Institute haben hierfür Routinen entwickelt.

Einfach, aber wirksam ist ein Prüfen der Häufigkeiten von „Keine Meinung“ und „Keine Antwort“. Sind diese Kategorien bei einigen Fragen häufig, so sollte das als Warnung dienen, die Eignung solcher Fragen noch einmal zu bedenken. Die Neigung in der kommerziellen Umfrageforschung, eine möglichst niedrige Zahl von Meinungslosen als Gütekriterium zu deuten, ist allerdings nicht akzeptabel. Es gibt je nach Thema schon einmal wirklich viele Meinungslose (Leverkus-Brüning 1966).

Besteht Grund zu der Vermutung, daß Fragen oder Fragengruppen Probleme bereiten, die weder im voraus entschieden noch mit Verlässlichkeit beim Interpretative probing ausreichend charakterisiert werden können, so bietet sich das Split-Ballot als Technik an. Das ist bekanntlich das Streuen der Interviews in zwei Versionen mit kontrolliert voneinander abweichenden Frageformulierungen. Man würde dies sicher sehr viel mehr als üblich benutzen, hätte das nicht zwei gravierende Nachteile: Es bringt Probleme mit sich bei der Verwaltung der Feldarbeit, und es verkleinert für viele Arten der Auswertung ein Sample um die Hälfte.³⁾

Das Programm der wissenschaftlichen Tagung sieht die Behandlung einiger spezialisierter Vorgehensweisen vor, wie Gruppendiskussionen. Sie brauchen hier nicht erörtert zu werden. Damit kann ich mich beschränken auf drei Vorgehensweisen, die sich allerdings für einen Einsatz als Routine nur begrenzt eignen.

Aus der Kognitionspsychologie – die hier von Bishop vorgestellt wird – ist eine Technik abzuleiten, die als „Laut denken“ (think out loud oder talk aloud) in die Literatur einging. Bei einigen Fragen – bei der betreffenden Untersuchung handelt es sich um ein Split-Ballot für elf Fragen mit Antwortvorgaben – werden die Antwortenden gebeten, alles laut auszusprechen, was ihnen in der Reaktion auf die Frage in den Sinn kommt. Dies wird dann anschließend bei der Bewertung der Antwortvorgaben hinzugezogen. Selbstverständlich könnte dies auch benutzt werden, um Kontexteffekte nachträglich abzuschätzen (Bishop 1992).

Das *Zufallsnachfassen* ist eigentlich keine Technik nur für den Pretest, sondern eignet sich ebenso für Haupterhebungen (Schuman 1966). In diesem Fall werden den Interviewern Zufallszahlen gegeben, aufgrund derer sie bei einer Untermenge der Fragen Nachfragen stellen sollen. Diese Zusatzantworten werden dann vom Interviewer nach einem Code mit Punkten versehen, welche die Übereinstimmung zwischen Fragezweck und Frageverständnis beim Befragten ausdrücken. In der betreffenden Untersuchung in Pakistan – Thema der Erhebung waren religiöse Überzeugungen und Praktiken – wurde so ermittelt, daß unter den etwa 20 % Antworten, welche der Frageabsicht nicht entsprachen, die Hälfte bei einem

3) Übrigens wird das Split-Ballot auch für einen gegenteiligen Zweck benutzt, nämlich zur Verlängerung des Fragenprogramms. Wenn das ganze Programm nicht in einer Erhebung unterzubringen ist, verteilt man die nicht in die vorgesehene Länge passenden weiteren Fragen auf zwei Hälften der Stichproben. George Gallup hat dieses Vorgehen öfters benutzt.

Fünftel des gesamten Fragenprogramms konzentriert war. Daraus ergab sich als praktische Folgerung, daß im nachhinein diese Fragen in der Auswertung getrennt berücksichtigt werden konnten.

Die in Deutschland wahrscheinlich informativste neuere empirische Untersuchung über Pretests wurde im Rahmen einer Arbeitsgruppe bei ZUMA im April 1989 durchgeführt. Die Darstellung erschien bisher nur als graue Literatur, so daß hier eine etwas ausführlichere Darstellung als bei den anderen durch Literatur zugänglichen Erhebungen angebracht ist. Forschungsgegenstand war das Nachfrageverhalten der Interviewer (Kreiselmaier, Prüfer, Rexroth 1989). Das Nachfragen (probing) kann selbstverständlich nur erfolgen, wenn der Interviewer erkennt, daß der Antwortende Probleme mit den ihm gestellten Fragen hat. In welchem Maße darf aber unterstellt werden, daß der Interviewer tatsächlich diese Probleme erkennt und darauf in angemessener Weise reagiert? Dieser Thematik wurde in einer methodischen Arbeit bei ZUMA anläßlich der Prüfung des Fragebogens für die Standarderhebung ALLBUS nachgegangen. Hauptthema des Fragebogens waren Wohnsituation und Wohnumwelt, aber dem Grundsatz nach handelte es sich um eine Mehr-Themen-Befragung.

Es wurde für den Zweck eines Pretest-Experiments kein eigens konstruierter Fragebogen benutzt, sondern das für die Haupterhebung erstellte Erhebungsinstrument. Die einzige wichtige Änderung gegenüber der sonst verwendeten Fassung bestand im Layout. Hier war neben den Fragen Raum für einen Code freigelassen worden, mit dem der Interviewer die Angemessenheit der Reaktion des Antwortenden auf eine Frage bewerten sollte:

- 1) = angemessen;
- 2) = problematisch;
- 3) = angemessen nach Rückfrage.

Wahrscheinlich wurde noch ein Code für das schließliche Fehlschlagen der Frage vergeben, aber das geht aus den Unterlagen nicht hervor. Als problematisch sollte das Verhalten des Befragten bewertet werden, wenn die Antwort nicht spontan und adäquat für den Stimulus erfolgte und wenn diese Abweichung als fragenspezifisch zu beurteilen war. Während des Interviews lief ein Tonband, dessen Übertragung dann mit den Eintragungen des Interviewers auf dem Fragebogen verglichen wurde. Diese Untersuchung hatte folgendes Ergebnis (siehe Übersicht 1, S. 24).

Übersicht 1

Individuelle Erkenntnisleistung bei inadäquaten Befragtenäußerungen

	Individuelle Erkenntnisleistung	Akademische Ausbildung	
	%	vorhanden	nicht vorhanden
Interviewer 1	22,1	-	x
Interviewer 2	37,3	-	x
Interviewer 3	51,4	x	-
Interviewer 4	63,4	x	-
Interviewer 5	81,3	x	-
Interviewer 6	86,0	x*	-
(* sozialwissenschaftliche Ausbildung)			

Die Autoren werteten das Ergebnis als ernüchternd (Kreiselmaier, Prüfer, Rexroth 1989). „Von allen im Pretest aufgetretenen inadäquaten Befragten-Äußerungen berichteten die Interviewer mit 48 % etwa die Hälfte“ (ibid., S. 22). Jeweils zur Hälfte bestand der Fehler des Interviewers darin, ein nach Tonband problematisches Verhalten des Befragten nicht zu erkennen oder zur anderen Hälfte in einer irrtümlichen Zuordnung. Auf seiten des Befragten wurden von den Autoren drei Hauptprobleme der Fragen gedeutet:

1. Ziel der Frage oder einzelne Begriffe wurden nicht verstanden;
2. es gab Schwierigkeiten mit den vorgegebenen Antwortkategorien und
3. in der Formulierung bediente man sich einer unangemessenen Sprachebene.

Einschränkend ist allerdings zu diesem sonst alarmierenden Befund zu sagen, daß die Interviewer über den Zweck der einzelnen Fragen nicht vorweg unterrichtet wurden. Die Autoren unterließen dies in der Absicht, experimentelle Bedingungen nicht zu verwässern, nämlich durch einen zusätzlichen Faktor „Verständnis des Interviewers“. So konnte der Interviewer auch nicht in dem normalen Maße seiner Aufgabe nachkommen, dem Befragten bei der Beantwortung von Fragen behilflich zu sein. Dieses Unterlassen ist kennzeichnend für eine unangemessene Orientierung bei einigen akademischen und darüber hinaus auch anderen nicht kommerziellen Umfrageinstituten. *„Es sei an dieser Stelle ausdrücklich darauf hingewiesen, daß die qualitative Pretestanalyse weniger ein ‚exakt wissenschaftliches Verfahren‘ ist als eine Art ‚Kunstlehre‘.“* (Porst 1985, S. 65.)

Die ZUMA-Untersuchung ist jedenfalls geeignet, noch einmal die Notwendigkeit von Pretests selbst bei erfahrenen Interviewern und bei wohlvertrauten Erhebungsthemen zu bekräftigen.

In diesem Zusammenhang sei noch abschließend auf die begrenzte Verlässlichkeit der Antworten auf einzelne Fragen zu verweisen. Aussagekräftig sind in der Umfrageforschung in erster Linie Antwortprofile. Dies ist das Ergebnis einer Gültigkeitsprüfung von Fragebatte-

rien (Fragen, die als Gruppe konzipiert sind zur Abdeckung eines Bereiches) des ALLBUS 1980. Der Koeffizient, mit dem die Gültigkeit ausgedrückt wird, ist Cronbachs Alpha, wobei 0 der schlechteste und 1 ein idealer Wert wäre. Dies ist das Ergebnis:

Übersicht 2
Reliabilitäten ausgewählter Itembatterien des ALLBUS 1980

Variablen	Itembatterien	Alpha
V 9 - V 15	Wichtigkeit von Lebensbereichen	.59474
V 16 - V 25	Arbeitsorientierungen	.85350
V 29 - V 37	Erziehungsziele	.84371
V 78 - V 83	Einstellungen zu Behörden	.71242
V 90 - V 100	Wahrnehmung von Konfliktgruppen	.83484
V 101 - V 104	Einstellungen zu Gastarbeitern	.76586
V 115 - V 118	Politische Issues	.28554
V 128 - V 131	Soziale Wünschbarkeit	.34170

„Die Reliabilitätskoeffizienten der Itembatterien ‚Politische Issues‘ und ‚Soziale Wünschbarkeit‘ müssen als sehr niedrig gelten. Dagegen sprechen Reliabilitätskoeffizienten von ca. .80, wie bei den Itembatterien ‚Arbeitsorientierungen‘; ‚Erziehungsziele‘, ‚Wahrnehmung von Konfliktgruppen‘ und ‚Einstellungen zu Gastarbeitern‘, für eine hohe Konsistenz des Antwortverhaltens hinsichtlich der Items dieser Batterien.“ (Porst 1985, S. 85.)

Abschließend ist noch vor einem Irrtum zu warnen, der immer wieder begangen wird, gerade wenn Mitarbeiter es nach den Standards eines Labor-Experiments besonders gut machen wollen. In solchen Fällen wird selbst für den Pretest dann eine möglichst große Annäherung an die Repräsentativität des Befragtenkreises angestrebt. Vorrang hat jedoch bei der Zusammenstellung des Personenkreises für einen Pretest ein anderes Kriterium. In diesem Stadium der Forschung ist es notwendig, als Teil des Befragtenkreises möglichst viele schwierige Fälle zu haben, selbst wenn sie in der Bevölkerung nicht so häufig sind. Werden in einer Erhebung etwa Wissensfragen unvermeidbar, so ist aus den Antworten die Angemessenheit des Frageniveaus offensichtlich bei weniger Gebildeten besser zu beurteilen. Beläßt das Kriterium Heterogenität des Befragtenkreises dann noch Raum für eine Annäherung an Repräsentativität – à la bonne heure (Galtung 1967, S. 138).

Literaturhinweise

Bishop, G. F. (1992): Qualitative Analysis of Question-Order and Context Effects: The Use of Think-Aloud Responses, in: N. Schwarz, S. Sudman, (Hrsg.): Context Effects in Social and Psychological Research, New York, S. 149 – 162.

Bohrnstedt, G., Mohler, P. Ph., Müller, W. (1987): An Empirical Study of the Reliability and Stability of Survey Research Items, in: Sociological Methods and Research, Jg. 15, Sonderheft, Nr. 3, Februar.

Cantril, H. (1951): Public Opinion 1935 – 1946, Princeton (New Jersey).

Converse, J. M., Presser, S. (o. J.): Survey Questions – Handcrafting the Standardized Questionnaire, in: Quantitative Applications in the Social Sciences, Sage Papers Nr. 63, Beverly Hills.

Diekmann, A. (1995): Empirische Sozialforschung, Reinbek.

Galtung, J. (1967): Theory and Methods of Social Research, Oslo, insbes. S. 137 – 142.

Jabine, T. B., Straf, M. L., Tanur, J. M., Tourangeau, R. (1984): Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines, Washington D. C.

Jahoda, M. (1981): To Publish or not to Publish, in: Journal of Social Issues, Jg. 37, Nr. 1, S. 208 – 220.

van Koolwijk, J. (1969): Unangenehme Fragen, in: Kölner Zeitschrift für Soziologie und Sozialpsychologie, Jg. 21, S. 864 – 875.

Kreiselmaier, J., Prüfer, P., Rexroth, M. (1989): Der Interviewer im Pretest, ZUMA Arbeitsbericht Nr. 89/14, Mannheim, April.

Kriz, J. (1981): Methodenkritik empirischer Sozialforschung, Stuttgart.

Kromrey, H. (1983): Empirische Sozialforschung, 2. Aufl., Opladen.

Leverkus-Brüning, I. (1966): Die Meinungslosen, in: G. Schmölders: Beiträge zur Verhaltensforschung, Heft 6, Berlin.

Loftus, E. (1986): Survey Remembering, in: Second Annual Research Conference, Washington, Bureau of the Census, Proceedings, March 23 – 26, S. 193 – 207.

Morton-Williams, J. (1993): Interviewer Approaches, Hants (England).

Oksenberg, L., Cannell, C., Kalton, G. (1991): New Strategies for Pretesting Questions, in: Journal of Official Statistics, Jg. 7, Nr. 3, S. 349 – 365.

Porst, P. (1985): Praxis der Umfrageforschung, Stuttgart.

Royston, P., Bercini, D., Sirken, M., Mingay, D. (1986): Questionnaire Design Research Laboratory, in: Proceedings of the Section on Survey Research, Washington D. C., S. 703 – 706.

Scheuch, E. K. (1953): Ein Interview über das Interview, Diplomarbeit, Köln.

Scheuch, E. K. (1973): Das Interview in der Sozialforschung, in: R. König, (Hrsg.), Handbuch der empirischen Sozialforschung, 3. Aufl., Bd. 2, Stuttgart, S. 66 – 190.

Schuman, H. (1966): The Random Probe – a Technique for Evaluating the Validity of Closed Questions, in: American Sociological Review, Jg. 1966, Nr. 2, April.

Sudman, S., Bradburn, N. M. (1983): Asking Questions. A Practical Guide to Questionnaire Design, 2. Auflage, San Francisco.

Willis, G., Royston, P., und Bercini, D. (1991): The Use of Verbal Report Methods in the Development and Testing of Survey Questionnaires, in: Applied Cognitive Psychology, Jg. 5, S. 251 – 267.

Der klassische Pretest und seine Techniken

Vorbemerkung

Praktische Erfahrungen werde ich bei meinen Ausführungen nicht ausblenden, ich werde mich aber auf die Ziele, Aufgaben und Techniken des Pretests konzentrieren. Dabei kommt manches zur Sprache, was sich der Forscher wünschen würde, aber die Praxis nur in begrenztem Maße zulässt. Wir werden in der anschließenden Diskussion vielleicht klären können, ob das nur für die kommerzielle Forschung zutrifft.

Illustrieren möchte ich die Ausführungen am Beispiel einer für uns sehr interessanten Forschungserfahrung: Eine große politische Partei wollte wissen, wie die Ziele und Inhalte ihrer Programmatik von ihrem Publikum gesehen und bewertet werden:

- Werden sie verstanden?
- Werden sie im intendierten Sinne verstanden?
- Welche Resonanz finden sie?

Diese Fragen wurden im Rahmen eines Forschungsprogramms beantwortet, das qualitative und quantitative Methoden der empirischen Sozialforschung miteinander verknüpfte.

Ich komme später darauf zurück.

1 Ziele und Aufgaben des Pretests

Lassen Sie mich zunächst kurz die wichtigsten Ziele und Aufgaben des Pretests skizzieren.

1. Der Pretest soll klären, welche *Akzeptanz* die Untersuchung bei den Zielpersonen findet.
 - Lösen die Untersuchungsziele, das Untersuchungsthema bzw. einzelne Untersuchungsthemen Aufgeschlossenheit, Desinteresse oder sogar Abwehr aus?
 - Erreichen die entwickelten Instrumente zur Herstellung der Beteiligungsbereitschaft (z. B. Begleitschreiben, telefonische Vorankündigungen) ihr Motivationsziel?

Die Rückmeldung aus dem Pretest soll also dem Forschungsteam Hinweise geben, ob die bisherigen Überlegungen zur Sicherung der Akzeptanz die erwünschte Wirkung haben oder überdacht werden müssen.

2. Der Pretest soll die *Erhebungssituation* explorieren.
 - Wieviel Zeit nimmt das Interview bzw. die Beobachtung oder Diskussion in Anspruch?
 - Welche Faktoren beeinflussen das Zeitbudget positiv oder negativ?

*) Horst Becker, polis, Gesellschaft für Politik- und Sozialforschung mbH, München.

- Entspricht die Erhebungssituation der intendierten Form oder ist mit unerwünschten Einflüssen zu rechnen? Zieht z. B. ein Arbeitnehmer bei einer Mitarbeiterbefragung den Betriebsrat hinzu?

Die Exploration der Erhebungssituation ist insbesondere bei schriftlichen Befragungen wichtig.

Sieht das Untersuchungsziel eine gemeinsame Bearbeitung des Fragebogens vor, z. B. von Haushaltsmitgliedern (bei einer Vollerhebung im Haushalt), dann muß der Pretest Hinweise dazu liefern, wie dieses Problem gelöst werden kann.

Wird eine möglichst von äußeren Einflüssen freie Bearbeitungssituation angestrebt, dann muß im Pretest geklärt werden, ob und wie das vom Forschungsteam entwickelte Verfahren realisiert werden kann.

Unerläßlich erscheint uns ein Pretest der Erhebungssituation mit Blick auf die beim Interview eingesetzten *neuen Techniken*. CAPI (Computer-assisted personal interviewing) bedeutet nicht nur eine neue Interviewtechnik, sondern auch eine neue Form des Dialogs mit den Befragungspersonen. Der zufriedene Hinweis, daß „es funktioniert“, gibt noch keine Antwort auf die Frage nach der Wirkung bei der Befragungsperson.

Durch die Exploration der Erhebungssituation kann das Forschungsteam wichtige Hinweise dazu erhalten, welche unerwünschten Einflüsse im Vorfeld der Untersuchung und im weiteren Verlauf durch eine präzise Instruktion der Interviewer bzw. Interviewerinnen vermieden werden müssen.

3. Der Pretest soll die *Erreichbarkeit der Zielpersonen* überprüfen. Dies ist insbesondere dann erforderlich, wenn sich die Untersuchung an Personen einer Zielgruppe richtet,
 - deren Lebensumstände dem Forschungsteam nicht ausreichend vertraut sind;
 - die z. B. aus Zeitknappheit äußerst schwer erreichbar sind;
 - deren Interviewbereitschaft von der Genehmigung anderer abhängig ist.

Durch den Pretest können somit z. B. die vom Forschungsteam entwickelten „Zugangsstrategien“ überprüft werden.

4. Der Pretest soll das entwickelte *Erhebungsinstrument* überprüfen.
 - Werden die Fragen von den Zielpersonen verstanden?
 - Gibt es ein gemeinsames Verständnis über die Inhalte und die Bedeutung der verwendeten Begriffe?
 - Ist die Reihenfolge der Themen bzw. der Fragen richtig gewählt? Stimmt der Spannungsbogen bzw. die Dramaturgie des Instruments?
 - Gibt es Themen oder Fragen, die in der für das Erhebungsinstrument entwickelten Form bei der Zielperson Abwehr oder auch den Abbruch des Interviews herbeiführen können?

- Verfügen die Zielpersonen über genügend Informationen, um die Fragen zu beantworten, oder – im Falle von Gruppendiskussionen – um über ein Thema zu diskutieren?
- Generell: Lassen sich mit dem Instrument die Forschungsfragen beantworten bzw. die Hypothesen prüfen?

Es ist heute vormittag bereits deutlich geworden: Im Zentrum des Interesses und auch der Pretest-Praxis steht der Test des Erhebungsinstruments. Deshalb werde ich mich im folgenden auch vor allem auf diese Aufgabe konzentrieren.

Lassen Sie mich aber zuvor noch eine weitere Aufgabe des Pretests nennen, die meines Wissens wissenschaftlich noch kaum untersucht wurde, in der Praxis aber eine erhebliche Bedeutung hat: Die Funktion des Pretests als *Instrument des Konflikt-Managements* zwischen denjenigen, die sehr viel wissen wollen und jenen, die diesen Wissensdurst in methodisch vertretbaren Grenzen halten müssen.

Die Situation ist allen hier Versammelten vertraut:

In der ersten Phase der Fragebogenentwicklung gemeinsam mit dem Auftraggeber entsteht häufig ein überdimensioniertes Programm, das die Befragungsperson eineinhalb bis zwei Stunden in Atem halten würde.

Von den Kunden wird jede Frage für unverzichtbar erklärt, man sei ohnehin schon an die nicht mehr zu unterschreitende Grenze des eigenen inhaltlichen Anspruchs gegangen.

In dieser Situation gibt es im Prinzip zwei bewährte Lösungsstrategien:

1. Die Forscher verzichten auf einen Teil ihrer Ansprüche an die spätere Datenanalyse und platzieren einige Themenblöcke auf verschiedene Fragebogensplits für Teilgesamtheiten.
2. Der Nachweis der Überforderung der Befragten wird durch den Pretest erbracht und aus dem Verlauf des Pretests ein Kürzungsvorschlag entwickelt.

Wir praktizieren gerne eine modifizierte Variante der zweiten Lösung:

Zunächst wird der Fragebogenumfang in seinem Zeitbudget von versierten Experten realistisch geschätzt. Ist das Programm eindeutig zu umfangreich, bitten wir die Kunden, unter dem Zwang zur Kürzung jede Frage skaliert zu bewerten:

- 3 = unverzichtbar,
- 2 = sehr wichtig,
- 1 = wünschenswert.

Alle bisherigen Erfahrungen führten zu dem Ergebnis, daß ein praktikabler Fragebogen bereits für den Pretest auf die „wünschenswerten Fragen“ (Kategorie 1) verzichten muß und nach dem Pretest auch manche „sehr wichtige“ Frage entbehrlich erscheint.

Ich habe vorhin gesagt: Das vorrangige Ziel von Pretests in der Forschungspraxis ist der Test des Erhebungsinstruments.

Im „klassischen“ Fall findet der Pretest einmalig kurz vor der eigentlichen Haupterhebung, quasi als „Simulation des Ernstfalles“ statt. Obwohl es kein einheitliches Vorgehen gibt, besteht die Übereinkunft der Sozialwissenschaft darin, daß ein Pretest unter möglichst realistischen Bedingungen, das heißt den Bedingungen der Hauptstudie, durchzuführen ist.

Seine zentrale Bedeutung bezieht der Pretest aus dem Mangel an konkreten Richtlinien zur Konstruktion von Fragen. Zwar gibt es eine Reihe von Rezepten und Empfehlungen sowie wichtige Einzeluntersuchungen über Effekte unterschiedlicher Frageformulierungen, der Reihenfolge von Fragestellungen sowie beispielsweise unterschiedlicher Skalen und ihrer Verbalisierung usw.

Dennoch bleibt auch bei gewissenhaftem Vorgehen und großer Erfahrung ein schwer kalkulierbares Restrisiko. Deshalb besteht die Hauptaufgabe eines Pretests darin, dieses Restrisiko zu minimieren und Hinweise auf Fehler und Mängel zu geben, die dann für die Haupterhebung beseitigt werden können.

Ich habe bisher über Ziele und Aufgaben des Pretests gesprochen und komme nun zu den „Techniken“, die beim Pretest angewandt werden.

2 Pretest-Techniken

Erstaunlicherweise finden sich in der Literatur relativ wenige Hinweise zum Methoden-Handwerk des Pretests. Diese Tatsache nährt den Verdacht (den auch die eigenen Erfahrungen bestätigen), daß in der Forschungspraxis nicht sehr häufig die Gelegenheit entsteht, mit ausreichenden finanziellen Mitteln und mit einem ausreichenden Zeitbudget einen aufgaben- und methodenadäquaten Pretest durchzuführen. Sehr häufig muß sich die Forschung mit einem Kompromiß an der Minimallinie zufriedengeben.

Zu dieser Minimallinie gehört, daß man den Pretest unter den Rahmenbedingungen der Hauptuntersuchung durchführt. Für die Durchführung unter möglichst realistischen Bedingungen werden erfahrene Interviewer herangezogen. In der Regel sind 20 bis 40 Interviews durchzuführen.

Die pretestrelevanten Hinweise werden dabei ausschließlich aus den Reaktionen der Befragten auf den Fragenstimulus abgeleitet. Die Interviewer sind somit mehr oder weniger „passive“ Informationssammler, die ohne Nachfragen „nur“ die Aufgabe haben, Probleme der Befragten bei der Beantwortung der einzelnen Fragen zu erkennen und an die Forscher weiterzumelden.

Zur Erfassung der Hinweise wird für jedes durchgeführte Interview im Anschluß an die Befragung ein Erfahrungsbericht in Form eines kleinen Fragebogens ausgefüllt:

Durch ihn wird erfaßt:

- Die Länge des jeweiligen Interviews.
- Die Anwesenheit und Eingriffe Dritter.
- Fragen zum allgemeinen Interviewverlauf, z. B. die allgemeine Antwortbereitschaft der Befragten, der Schwierigkeitsgrad für den Interviewer und den Befragten. Die Bewertung erfolgt dabei mit Hilfe von Einstufungsskalen.

- Allgemeine Eindrücke des Interviewers bei der Durchführung sowie die Beurteilung des Spannungsbogens. Erfasst wird dabei die Zahl bzw. die Gefahr von Abbrüchen sowie die Passagen im Fragebogen, bei denen die Zielperson unruhig und/oder unkonzentriert wurde.
- Hinweise auf unvollständige oder nicht eindeutige Antwortkategorien.
- Hinweise auf unverständliche Begriffe oder Formulierungen.
- Probleme und Schwierigkeiten bei einzelnen Fragen.

Trotz der hilfreichen Hinweise, die man durch ein solches Vorgehen erhält, lassen sich auch zwei gravierende Nachteile bei der Durchführung des Pretests anführen.

1. Das „passive“ Informationssammeln, zu dem die Interviewer verpflichtet werden, läßt keine Rückkopplung auf das individuelle Fragenverständnis zu.
2. Werden die Hinweise quasi als Gedächtnisprotokoll von den Interviewern nach den Befragungen notiert, ist man somit ausschließlich auf die Beobachtungsgabe und das inhaltliche Verständnis der Interviewer angewiesen.

Obwohl mit dem Erfahrungsbericht der Versuch einer Systematisierung der Beobachtungen unternommen wird, liegen die Schwächen eines solchen Verfahrens auf der Hand.

Die – wie ich meine – interessante Untersuchung zum Thema „Der Interviewer im Pretest“ von Kreiselmaier, Prüfer und Rexroth¹⁾ gibt Hinweise darauf, wie hoch die „Fehlerquote“ bei einem solchen Pretest sein kann.

Von den pretestrelevanten Hinweisen, die die Autoren mit Hilfe der auf Tonband aufgenommenen Interviews nachträglich ermittelten, wurden von den eingesetzten Interviewern nur 48 % erfaßt.

Rund jeder zweite pretestrelevante Hinweis wurde von den Interviewern demnach *nicht* identifiziert.

Kreiselmaier, Prüfer und Rexroth geben auf der Grundlage ihrer Studie einige Hinweise auf Techniken, die ich hier aufgreifen möchte.

Die Autoren betonen den prozeßhaften Charakter der Fragebogenentwicklung und fordern einen zweistufigen Pretest.

In der ersten Phase, der sogenannten *Pilotphase*, kommen in einem bereits standardisierten Entwurf eines Fragebogens vor allem „aktive“ Verfahren zum Einsatz. Durchgeführt wird dieser Pretest von den Experten bzw. Projektleitern, die die Zielperson über den Testcharakter informieren und so auch ihr einen aktiven Part zuweisen. Der standardisierte Entwurf des Fragebogens umfaßt dabei zusätzlich offene bis standardisierte Nachfaßfragen zum Verständnis der Fragen sowie der einzelnen Begriffe.

Auf Grundlage dieser Pilotphase wird der Fragebogen überarbeitet und geht dann in die sogenannte *Testphase*.

1) Kreiselmaier, J., Prüfer, P., Rexroth, M.: Der Interviewer im Pretest. Evaluation der Interviewerleistung und Entwurf eines neuen Pretestkonzepts, ZUMA-Arbeitsbericht Nr. 89/14, Mannheim 1989.

Dieser zweite Pretest simuliert die Bedingungen der Hauptstudie. Eingesetzt werden Interviewer, die mit Hilfe passiver Verfahren befragten.

Ausgehend von diesen Erfahrungen zur Rolle der Interviewer im Pretest ist anzuraten, diese Testphase mit Tonbandmitschnitten zu begleiten und so die Mitarbeit von Projektleitern weiter zu intensivieren.

Eine Verbesserung der Interviewerleistung kann auch durch eine entsprechende Schulung erreicht werden, die die Aufmerksamkeit für die begleitende und systematische Bewertung der Fragen schärft.

Eine begleitende Bewertung kann erfolgen, indem jede Frage mit Codeziffern versehen wird, wie es Kreiselmaier u. a. in ihrer Untersuchung praktizieren.

Mit einer numerischen Bewertung, z. B. 1 für adäquates und 2 für inadäquates Antworten, errechnet sich bei jeder Frage für alle realisierten Interviews ein Wert, der das Verhältnis der adäquaten zu den inadäquaten Antworten ausdrückt. Dieser Wert kann dann bereits als Hinweis auf die Güte einer Frage gelten bzw. systematische Hinweise auf besonders schwierige Fragen geben. Die Schulung der Interviewer müßte darauf zielen, ihre Beobachtungsfähigkeit zu schärfen. Eine Hilfe dabei wäre z. B., „adäquates“ Antwortverhalten zu definieren als spontanes und/oder im Sinne der Antwortvorgaben richtiges Antworten. Alles, was die Befragten zum Stocken, zur Nachfrage anregt oder im Sinne der Vorgaben zu falschen Antworten führt, wäre dagegen als „inadäquates“ Verhalten einzustufen.

Die Gründe für „richtiges“ oder „falsches“ Antworten würden sich allerdings weiterhin nur aus den Interviewermitschriften erschließen lassen. Der begleitende Erhebungsbogen hat also eine wichtige Bedeutung. Auch hier ließe sich mit gezielter Schulung vermeiden, daß die Rückmeldungen der Interviewer zu pauschal geraten.

3 Das Beispiel einer Programm-Untersuchung

Lassen Sie mich abschließend noch in aller Kürze von dem Pretest zu der Untersuchung berichten, die ich zu Beginn angesprochen habe, nämlich die Programmstudie für eine politische Partei.

Die programmatischen Aussagen bezogen sich auf eine breite Themenpalette, die von der äußeren und inneren Sicherheit über Wirtschafts- und Sozialpolitik bis zur Umweltpolitik reichte.

Die vorgegebenen politisch-programmatischen Aussagen mußten von einem abstrakten Niveau oder aus einer Spezialsprache in Forschungsfragen übersetzt werden, die auch politisch nicht besonders interessierte und informierte Bürgerinnen und Bürger verstehen.

Man kann sagen, daß das Programm bereits von unserem Forschungsteam einem kritischen Pretest unterzogen wurde.

Parallel dazu wurden von einem anderen Forschungsinstitut in leitfadengestützten Explorationen mit Zielpersonen aus unterschiedlichen Regionen (z. B. Ost/West), unterschiedlichen sozialen Schichten und Altersgruppen die Reaktion auf die wichtigsten Programmaussagen in Erfahrung gebracht.

Alle vorbereitenden Schritte mündeten dann in die Entwicklung des Fragebogens. Es ist leicht vorstellbar, daß es sich dabei – von einigen wahlsoziologischen Standardinstrumenten abgesehen – um ein völlig neu entwickeltes Instrument handelte.

Der Pretest wurde einerseits in der „klassischen Form“ durchgeführt: Die Interviewer des für die Feldarbeit zuständigen Instituts (Marplan) bewerteten jede Frage unter Aspekten der Verständlichkeit und der Eindeutigkeit, aber auch unter Aspekten der Überforderung. Diese Bewertungen wurden in Form von Pretest-Protokollen übermittelt.

Daneben wurden von den Projektleitern unseres Instituts Interviews geführt, in denen mit den Zielpersonen der ausführliche Dialog gesucht wurde. Dieser unmittelbare „sinnliche Eindruck“ erwies sich als äußerst nützlich, weil dabei inhaltliche und semantische Überforderungen oder auch Mißverständnisse eindrucksvoll aufgezeigt wurden. Es fehlt die Zeit, um die vielen nützlichen Lehren aus diesem Pretest im einzelnen zu schildern.

Deshalb nur soviel:

- Der Pretest führte zu einer Veränderung der *Fragebogen-Dramaturgie*, weil nun der unterschiedliche Grad des Interesses und der Aufmerksamkeit für die einzelnen Themenbereiche bekannt war und berücksichtigt werden konnte.
- Der Pretest machte deutlich, daß in der Untersuchung auf *bestimmte Themen verzichtet* werden kann, weil selbst mit dem besten Frageninstrument nichts anderes in Erfahrung gebracht werden kann als Verwirrung, zwiespältige oder auch Ohnmachtsgefühle. Als Prototyp sei die Rolle von UNO und NATO sowie der Bundeswehr mit Blick auf das ehemalige Jugoslawien oder andere internationale Krisengebiete erwähnt.
- Der Pretest zeigte auf, daß selbst solche *Begriffe der politischen Umgangssprache*, die Politiker und Medien täglich benutzen, nicht oder zum Teil auch falsch verstanden werden. Als Beispiele seien nur genannt: Geldwäsche, Kinderfreibetrag, Ehegattensplitting oder Fehlbelegungsabgabe.
- Durch Fragebogen-Splits beim Pretest wurde das Antwortverhalten bei unterschiedlicher Erläuterung von Programmzielen getestet, etwa bei einem so komplexen Thema wie der Ökosteuer. Die Ergebnisse des Pretests führten zur Entwicklung eines Frageninstruments, das schrittweise von grundsätzlichen zu inhaltlich alternativ formulierten Angeboten führte.

Im Ergebnis erwies sich der Pretest für dieses Projekt als unerläßliches Instrument zur Erarbeitung eines möglichst punktgenauen Fragenprogramms und zur Herstellung einer adäquaten Interviewsituation.

Wie so oft konnte das wünschenswerte Vorgehen nur begrenzt realisiert werden. Es mußte durch das Wissen ergänzt werden, das in langjähriger Forschungsarbeit entsteht.

Abschließend noch eine Bemerkung:

Natürlich sollte durch die spätere Hauptuntersuchung nicht ermittelt werden, was der Auftraggeber tun oder lassen soll. Die Untersuchung gab aber sicherlich wichtige Hinweise zur Kommunikation der Programmziele in der Öffentlichkeit.

Bewertung und Beurteilung von Fragebogen durch Gruppendiskussionen

1 Gruppendiskussionen in der Sozialforschung

Gruppendiskussionen wurden in der qualitativen Marktforschung bisher als Pretest zur ersten Sammlung von Meinungen, Vorstellungen, Wünschen, Erwartungen, Befürchtungen im Hinblick auf einen Meinungsgegenstand verstanden.

In diesem Sinne dienen Gruppendiskussionen nicht als Pretest eines Fragebogens, es werden also nicht die Fragen zur Diskussion gestellt, sondern es handelt sich um eine Materialsammlung zur Formulierung eines Fragebogens.

Wenn dies üblicherweise so ist, bedeutet dies natürlich nicht, daß man nicht auch Fragen zur Diskussion stellen könnte und kann. Dies ist eben nur ungewöhnlich!

Die Gruppendiskussion als Instrument der Materialsammlung zur Formulierung eines Fragebogens stellt im Rahmen eines Forschungsprogramms einen ganz „normalen“ Forschungsschritt dar. Er wird auch meist nicht aufgrund der dafür erforderlichen Kosten abgelehnt (zur Zeit etwa zwischen DM 6 000,- und DM 8 000,- für eine Testgruppe mit „normaler“ Zielgruppe).

Die Gruppendiskussion als Pretest und Materialsammlung ermöglicht natürlich auch das Hinterfragen von möglicherweise schwer verständlichen Begriffen, von Statements, also nicht nur der Formulierung von Fragen.

Die Gruppendiskussion ersetzt allerdings *nicht* einen Pretest des dann vorliegenden Fragebogens als solchen.

Gruppendiskussionen werden in der Marktforschung allerdings zunehmend nicht mehr ausschließlich als Pretest eingesetzt.

Sie dienen als (scheinbar) kostengünstiges Verfahren zur schnellen – und meist oberflächlichen – Analyse und Beurteilung von Zeitschriften, Zeitschriftentiteln und -inhalten oder auch zur Ermittlung der Wirkungsweise von Produkten, Packungen, Werbemitteln, Kauf- und Verwendungsmotiven etc.

Daß dabei gelegentlich nur zufällige – von der Zusammensetzung der Gruppe bestimmte – Ergebnisse ermittelt werden, wird dabei ganz offensichtlich in Kauf genommen. Es ist deshalb ganz wesentlich, zunächst einmal die *Zusammensetzung* der Gruppen, also die Auswahl der Befragungspersonen zu beachten. Weiß man allerdings keine Details über die Verteilung, z. B. die Akzeptanz des Untersuchungsgegenstandes am Markt oder die Verteilung der Meinungen, so kann es schnell zu falschen Schlußfolgerungen kommen. Will man dagegen nur Vorstellungen über einen Meinungsgegenstand ermitteln, um diese nachher möglichst „volksnah“, also im Sprachgebrauch der Zielgruppe zu formulieren, so ist die Gefahr geringer, aber immer noch existent.

*) Dr. Klaus Haupt, psyma, arbeitsgruppe für psychologische marktanalysen GmbH, Rückersdorf.

Es ist dann letztlich eine Kostenfrage, welche Anzahl von Gruppendiskussionen man sich leisten kann, um die unterschiedlichen Teile eines Fragebogens zu entwickeln und hinsichtlich z. B. ihres semantischen Gehalts zu überprüfen.

Der Boom dieser Art der Nutzung von Gruppendiskussionen ist aus USA und England zu uns „übergeschwappt“, wo aufgrund wissenschaftlicher Traditionen – im übrigen trotz eines Ernst Dichter – die Einzelexploration weniger selbstverständlich ist. Meist begnügt man sich mit zwei Gruppendiskussionen und hält sie für ausreichend aussagekräftig für Deutschland. Neuerdings wird offensichtlich sogar „qualitative Forschung“ mit Gruppendiskussionen gleichgesetzt.

Geht man davon aus, daß eine Gruppendiskussion durchschnittlich zwischen DM 6 000,- und DM 8 000,- (bei normal auffindbaren Teilnehmern) kostet, dann frage ich mich, ob man für das Geld von zwei Gruppendiskussionen nicht etwa 50 ausführliche Einzelexplorationen in Auftrag geben sollte und davon wesentlich mehr und informativere Aussagen bekommen könnte, im übrigen auch Aussagen zur Formulierung und Überprüfung von Fragebogen.

Es werden dann nämlich 50 Einzelnennungen und soziodemografische oder psychografische Merkmale getrennt erhoben. Zwei Gruppendiskussionen mit zusammen 16 bis 18 Teilnehmern bringen dagegen erfahrungsgemäß maximal 8 bis 10 verschiedene Meinungen, bei denen die Häufigkeit ihres Vorkommens im Gegensatz zur Exploration von 50 Einzelpersonen nicht festgestellt werden kann.

Gruppendiskussionen sollte man also möglichst nur dann als Pretest einsetzen, wenn man absolut keine Vorstellungen vom Meinungsgegenstand hat. Dies allerdings unter der Voraussetzung, daß die Gesprächsführung in beiden Fällen lediglich anhand eines Leitfadens und nicht eines Fragebogens erfolgt.

Aber es ist der Vorteil von Gruppendiskussionen – natürlich auch als Pretest – daß sie in kurzer Zeit durchführbar, ausgewertet, diskutiert und präsentiert sind. Einzelexplorationen benötigen dagegen länger von der Feldarbeit bis zur Präsentation, liefern aber eben sehr viel detailliertere Ergebnisse.

Weil bei Gruppendiskussionen der Auftraggeber – hinter dem Spiegel – teilnehmen kann, erfüllen sie eine wesentliche zusätzliche Funktion: Die Produkt-, Marketing- und Werbemanager, aber auch der Sozialforscher am Schreibtisch, können sich auf diese Weise selbst einen Einblick verschaffen, wie die Menschen, denen sie etwas verkaufen wollen oder deren Auskünfte sie für ihre Forschungstätigkeit brauchen, wirklich aussehen, wie sie sich in der Gruppe verhalten, ob und wie sie sich mit dem Produkt oder dem Thema zu identifizieren oder auseinanderzusetzen bereit sind. Auch die Mimik und Gestik und das Sozialverhalten der Teilnehmer läßt dabei wichtige Rückschlüsse zu.

Dieses hautnahe Erleben menschlicher Reaktionsweisen von Verbrauchern sollten sich Industriemanager regelmäßig gönnen, und auch Sozialforschern tut es gut, gelegentlich zu sehen, wie die Menschen wirklich aussehen:

So formulieren sie doch meist lediglich den Fragebogen und beschäftigen sich dann im allgemeinen erst wieder mit den Rohtabellen, selten mit den ausgefüllten Fragebogen, in denen sie zumindest wörtliche Äußerungen nachlesen könnten.

Wie Menschen aussehen, sich verhalten, welche Erwartungen, Wünsche, Hoffnungen sie haben, wissen sie aus den Ergebnissen ihrer Studien, im allgemeinen aber nicht aus der Wirklichkeit. Der Wirklichkeit der „normalen“ Menschen verschiedener Altersklassen und sozialer Schichten begegnen sie selten z. B. auf dem Weg zu ihrer Arbeitsstätte oder auch sonst, weil sie nicht unbedingt zur gleichen Zeit unterwegs sind, keine öffentlichen Verkehrsmittel benutzen, einen anderen Freundeskreis haben usw.

Insofern ist der „Gruppendiskussions-Tourismus“ nicht nur entschuldbar, sondern sollte für alle Forscher sogar zum „must“ werden. Man sollte wissen, wie die Menschen aussehen, deren Äußerungen man dann weiterverarbeitet, wie sie sich anziehen, welches Selbstverständnis, welche Hemmungen und Probleme sie haben, wie sie sich verhalten, worauf sie stolz sind und wovon sie reden und was sie verschweigen.

Gerade auch für Forschungsaufgaben im und für den öffentlichen Bereich und die Politik und die Themen, über die die Menschen zu Hause und mit Freunden reden (z. B. die wirtschaftliche Entwicklung, die Arbeitslosigkeit, die Notwendigkeit und Verwendung von Abgaben und Steuern, die Pflegeversicherung etc.), wäre es – schon um die Staats- und Politikerverdrossenheit abzubauen – von großer Bedeutung, Gruppendiskussionen nicht nur als Pretest, sondern zur Ermittlung von Stimmungsbildern und dem unmittelbaren Erleben eines Ausschnitts der Wirklichkeit durchzuführen! Aber die Gruppen sollten dann „richtig“ zusammengesetzt und der Diskussionsleiter erfahren und gut „gebrieft“ sein.

Nach diesem Ausflug in eine zusätzliche Funktion (nämlich in die Wiedergewinnung eines Realitätsbezugs) von Gruppendiskussionen in der Markt- und Sozialforschung kommen wir zur Gruppendiskussion als Pretest zur Entwicklung von Fragebogen.

Natürlich ist in der Sozialforschung, aber auch bei bestimmten Aufgaben der Marktforschung die Gruppendiskussion ein Pretest-Instrument. Sie dient der Sammlung erster mit dem Untersuchungsgegenstand verbundener Informationen, Meinungen, Vorstellungen, Erwartungen, Befürchtungen, Wünsche, Hoffnungen etc.

Man nutzt Gruppendiskussionen zum besseren Verständnis von Aufgabenstellungen und zur ersten Information.

2 Struktur von Gruppendiskussionen

Die *inhaltliche und technische Vorbereitung* von Gruppendiskussionen ist wichtig.

Sie umfaßt

- die Formulierung und Übermittlung der Aufgabenstellung an den Diskussionsleiter in Form eines ausführlichen mündlichen Briefings, möglichst mit Brainstormingcharakter – also in Form einer Gruppendiskussion unter Fachleuten,
- die Definition der Anforderungen an den zu entwickelnden Fragebogen und seinen Aufbau bzw. seine Struktur,
- die Formulierung eines Themenkatalogs für die Gruppendiskussion, der dann zur Grundlage für den Diskussionsleitfaden wird,
- die Diskussion möglicher Alternativen zur Gruppendiskussion (braucht das Thema unbedingt eine Gruppendiskussion als Pretest?),

- die Formulierung eines Leitfadens für die Gruppendiskussion und Abstimmung des Leitfadens mit dem Auftraggeber.

Dazu im einzelnen folgendes:

2.1 Vorbereitung der Gruppendiskussion

Die Formulierung und Übermittlung der *Aufgabenstellung* in Form eines ausführlichen Briefings:

- Was soll ermittelt werden, was ist das Ziel der Untersuchung ?
- Welche Maßnahmen, Konsequenzen sollen aus den Ergebnissen gezogen werden ?

2.2 Die Definition der Anforderungen an den zu entwickelnden Fragebogen

- Dauer der Befragung.
- Offene Fragen:

Typus der offenen Fragen:

- Verhaltens- /Meinungsfragen,
 - freie Assoziationen,
 - projektive Fragen,
 - Satzergänzungen.
- Geschlossene Fragen:
Typus der geschlossenen Fragen:
 - Ja/nein-Fragen,
 - formulierte Antwortvorgaben,
 - Statementbatterien.
 - Auswertungskriterien für den Fragebogen, die möglicherweise Abfolge und Fragenformulierung beeinflussen.

Während der Gruppendiskussionen ist auf eine möglichst vollständige Erhebung aller möglichen Dimensionen zu achten, die im Zusammenhang mit dem Untersuchungsgegenstand relevant sind. Um dieses Ziel zu erreichen, ist die Qualifikation des Diskussionsleiters sehr wichtig. Wesentlich ist, daß er weder suggestiv noch determinierend (im Hinblick auf gewünschte Antworten!) die Diskussion leitet.

Bei der *Auswertung der Gruppendiskussion* sind die für den Untersuchungsgegenstand relevanten Dimensionen zu ermitteln und entsprechend für den Fragebogen zu formulieren.

2.3 Formulierung eines Themenkatalogs für die Gruppendiskussion

2.3.1 Brainstorming der Fachleute, Hypothesen

- Welche verschiedenen Dimensionen und Aspekte könnten eine Rolle spielen (zentrale Aspekte, Randaspekte) ?
- Welche Informationen, Kenntnisse, Erfahrungen existieren und müssen nur auf ihre Relevanz überprüft/hinterfragt werden ?
- Formulierung von Hypothesen.
- Bedeutung und Aktualität der Thematik für den Befragten.
- Nähe zur Intimsphäre/ zu Tabus (generell oder bei speziellen Zielgruppen).
- Möglicherweise beim Befragten durch das Thema bedingte Schwierigkeiten bezüglich des Verständnisses.
- Definition von Begriffen, semantische Probleme, Verständnisfragen.
- Vordergründigkeit/Tiefe. (Gibt es „schnelle“ Antworten, mit denen das Thema „abgefertigt“ werden kann ?)
- Anteil eigener Erfahrungen versus Images vom Untersuchungsgegenstand. (Sind Erfahrungen/Fakten oder Meinungen/Images als Antwort zu erwarten, und wie kommt man an die dahinter liegenden Motivstrukturen ?)

2.3.2 Desk-Research

- Sammlung vorhandener Fakten und Daten, Meinungsverteilungen.
- Literatur, Meinungen und Erfahrungen anderer mit dem Thema.

2.4 Die Diskussion möglicher Alternativen zu Gruppendiskussionen

Z. B. Einzelexplorationen:

- | | |
|------------------------|--|
| <i>Vorteile</i> | <ul style="list-style-type: none">– Zum gleichen Preis größere Zahl von Auskunftspersonen.– Unbeeinflusste Aussagen, keine Gruppeneffekte.– Größere Zahl wörtlicher Aussagen, eigenständige Formulierungen (die später als Vorgaben besser geeignet sind).– Erste Gewichtung der ermittelten Argumente und Dimensionen. |
| <i>Übereinstimmung</i> | <ul style="list-style-type: none">– Leitfadengespräche wie bei der Gruppe.– Gleiche Vorbereitung.– Gleiche Qualifikation wie der Diskussionsleiter. |
| <i>Nachteile</i> | <ul style="list-style-type: none">– Größerer Zeitbedarf.– Erhöhter Auswertungsaufwand.– Keine gruppendynamischen Prozesse. |

2.5 Formulierung eines Leitfadens für die Gruppendiskussion

Berücksichtigung aller vorgenannten Gesichtspunkte.

3 Durchführung und Ablauf von Gruppendiskussionen

Zur Vorbereitung und Durchführung von Gruppendiskussionen bedarf es folgender *Qualifikationen des Diskussionsleiters*:

- optimale Ausbildung: Psychologe/Soziologe/Sozialforscher;
- gut „gebrieft“, mit Aufgabenstellung vertraut;
- erfahren in der Diskussionsleitung, das bedeutet:
 - vermeidet Konsensbildung der Gruppe, hält Gespräch in Gang;
 - denkt in jeder Phase mit, ist selbst assoziativ, provoziert, reizt zum Widerspruch, ohne suggestiv und determinierend zu sein;
 - ist nicht an Abfolge des Leitfadens gebunden, hat das Thema und damit den Leitfaden im Kopf und kann auf Anregungen aus der Gruppe eingehen, Themen aufgreifen und Abschweifungen abwürgen;
 - fördert Stille, dämpft Laute, erkennt Meinungsbildner und deren Einfluß auf die Dynamik der Gruppe;
 - wirkt sympathisch, offen, einfühlsam; spricht stets kontrolliert;
 - versteht Mimik, Gestik, Nicht-Gesagtes und verwertet dies bei seinen Fragestellungen.

Daraus ergeben sich gleichzeitig die spezifischen Möglichkeiten der Gruppendiskussionen¹⁾ und die weiteren Anforderungen an den Diskussionsleiter:

- Erfassen des Individualverhaltens und gruppendynamischer Prozesse.
- Stimulierung der Teilnehmer zu mehr Kreativität/Phantasie und Aktualisierung tieferliegender Einstellungen und Motive. Es können weitergehende Äußerungen als nur Primär-Assoziationen (Haus – Tür) festgestellt werden.
- Schnellerer Abbau der im Interview häufig zu beobachtenden Befragungssituation; Hemmungen, Ängste, Widerstände werden abgebaut; die Teilnehmer lassen sich provozieren.
- Der Einfluß von Meinungsbildungsprozessen auf die Bildung späterer Verhaltens- und Einstellungsweisen wird sichtbar.
- Die Simulation von Situationen während der Gruppendiskussion ähnelt dem normalen Kommunikations- und Meinungsbildungsprozeß mehr als bei der Einzelexploration.
- Vermeiden eines Abgleitens in Stammtisch- und Kaffeekränzchen-Gespräche.

1) Nach Salcher, Ernst, F.: Psychologische Marktforschung, Berlin, New York, 1978, S. 54.

Insgesamt fördert die unmittelbare Reaktion auf Meinungen und Vorstellungen anderer das Nachdenken über die eigene Position. Dies führt zu einer vertieften Auseinandersetzung mit dem Meinungsgegenstand. Dieser Prozeß findet sowohl innerhalb der Gruppe als auch zwischen Gruppe und Diskussionsleiter statt.

Der Diskussionsleiter hat sich auf die jeweiligen Teilnehmer an der Gruppendiskussion einzustellen. Er muß sich ihrer Sprache anpassen, indem er ihre Semantik übernimmt, möglichst nicht suggestiv fragt – trotzdem provoziert, um den Gruppenprozeß in Gang zu bringen – und die Körpersprache registriert und für seine Leitungsaufgabe nutzt. Dabei muß er beachten, daß die Teilnehmer eine bestimmte Anzahl von Personen sind, die ein zeitlich begrenztes gemeinsames Anliegen zusammenführt. Sie haben sich bereit erklärt, an einer Diskussion teilzunehmen, ohne die anderen Teilnehmer, ohne das Thema und ohne die Umgebung, in der sie stattfindet, zu kennen. Damit wird eine gewisse Vorauswahl getroffen.

Für die Formulierung von Fragebogen (Fragen, Antwortvorgaben, Statements) sind die wörtlichen Aussagen, die Art und Weise, wie ein Gegenstand oder Sachverhalt beschrieben wird, von besonderer Bedeutung. Es können geplante Fragenformulierungen sowohl auf ihre Verständlichkeit unmittelbar ebenso überprüft werden (Diskussion der Formulierung) als auch Formulierungen der Befragten übernommen werden als natürlich auch neue Dimensionen für den Fragebogen gefunden werden.

Es stellt sich deshalb im Zusammenhang mit unserem Thema die Frage, ob sich ein Pretest an einem bereits (vor-)formulierten Fragebogen orientieren soll oder ob die Fragebogenformulierung erst nach der Diskussion der Ergebnisse der Gruppendiskussion erfolgen sollte.

Wir empfehlen letzteres, allerdings unter der Voraussetzung eines sehr detaillierten Briefings und Themenkatalogs, wie bereits zuvor beschrieben.

3.1 Die Zusammensetzung der Gruppe

Wichtig ist, daß zu Gruppendiskussionen eingeladene Personen bestimmte, den Untersuchungsgegenstand betreffende Gemeinsamkeiten haben. Je seltener diese Gemeinsamkeit vorkommt – z. B. Zugehörigkeit zu einer seltenen Berufsgruppe, Verwendung eines bestimmten Produkts oder Vertreter einer Meinungsgruppe –, um so (kosten-)aufwendiger ist ihre „Beschaffung“.

Die Teilnehmer sollten mögliche unterschiedliche Dimensionen hinsichtlich des Untersuchungsgegenstandes aufgrund ihres soziografischen oder psychografischen Status widerspiegeln – natürlich kann dies nur in dem Maß sein, in dem die Meinungsverteilung bekannt ist.

Der sie einladende Sozialforscher kennt zunächst nur allein diese Gemeinsamkeiten. Der Teilnehmer dagegen ist bestrebt zu erfahren, worum es eigentlich geht und meist unsicher, ob er die in ihn gesetzten Erwartungen erfüllen kann.

Insofern existiert, wiewohl bei sich spontan bildenden Gruppen oder bei existierenden Gruppen, zunächst kein gemeinsames Anliegen, keine Verbindung durch gemeinsame Aktivitäten.

Diese fehlenden Merkmale bestehender Gruppen, nämlich

- eigene Identität,
- Wir-Gefühl,
- soziale Ordnung,
- Verhaltensnormen/emotionaler Kontakt sowie
- Wert und Ziele

müssen erst im Laufe der Diskussion durch den Diskussionsleiter aufgebaut bzw. vorgegeben werden. Die Art und Weise, wie sie sich entwickeln, kann bereits zum Untersuchungsergebnis gehören, denn das vorgegebene Thema kann zu unterschiedlicher Identifikations- und Auseinandersetzungsbereitschaft führen. Dies wiederum sollte zu einer entsprechenden Reaktion im Verlauf der Diskussion führen. Wie eine Ja/nein-Antwort ein Gespräch abrupt beendet, so tötet auch Konsens eine Diskussion. Zur Ausleuchtung aller möglichen Aspekte und Dimensionen eines Untersuchungsgegenstandes ist die kontroverse Diskussion von entscheidender Bedeutung.

Der Diskussionsleiter muß durch Vorgabe der zur Diskussion stehenden Untersuchungsgegenstände eine gewisse Identität und Identifikationsbereitschaft mit dem Thema und damit interaktive Beziehungen entwickeln sowie die Spielregeln (indem er „Dominante“ bremst und „Gehemmte“ ermuntert) und die Ziele vorgeben. Er muß also den gruppendynamischen Prozeß in Gang bringen.

Ob dies gelingt, hängt meist mehr vom Untersuchungsgegenstand als vom Diskussionsleiter oder den Teilnehmern ab. Die Gruppendynamik entwickelt sich aus der kontroversen Diskussion.

Das vorgegebene Thema kann allerdings sehr unterschiedlich strukturiert sein und z. B. einen sehr starken Ich-Bezug haben, der der Auseinandersetzung förderlich ist (Kinder, Autos, Urlaub, Gesundheit) oder nur von geringem Interesse sein.

Von geringem Interesse sind meist

- prestigearme, der Bedeutung der eigenen Person nur wenig förderliche Themen,
- Tabu-Themen/Privatsphäre; Themen von mangelnder Aktualität und geringem allgemeinen Interesse.

Insofern eignen sich nicht alle Themen für die Behandlung in Gruppendiskussionen bzw. erfordern zusätzliche auflockernde Einschübe, wie z. B. die schriftliche Beantwortung von Fragen und deren spätere Diskussion, Erarbeiten von Collagen, Rollenspiele, Assoziationen aus anderen Bereichen etc.

Allerdings dürfen sie nicht zu gleichzeitigem Abfragen von mehreren Personen degenerieren.

Assoziative und projektive Techniken der klinischen Psychologie eignen sich besonders, die Kreativität der Teilnehmer und damit gruppendynamische Prozesse anzuregen. (TAT²⁾-artige Geschichten, Beschreibung des Untersuchungsgegenstands als Mensch, Tier etc.)

Im Zusammenhang mit der Zusammensetzung der Gruppe ist auch das Thema der Homogenität/Heterogenität von Meinungen und damit von Befragten Gruppen zu diskutieren. Homogenität der Gruppe fördert die Konsensbildung. Dies kann man mit der Durchführung mehrerer aufeinander aufbauender, von unterschiedlichen Ansätzen bzw. Ausgangsthemen ausgehender bzw. auf den Ergebnissen der letzten Diskussionsrunde basierender Gruppendiskussionen verhindern.

Man kann Homogenität/Heterogenität, aber auch – zusätzlich oder grundsätzlich – dadurch erreichen, daß Personen mit zu erwartenden divergierenden Einstellungen zum Untersuchungsgegenstand eingeladen werden. So kann man Yellow-press-Zeitschriften untersuchen, indem man in einer Gruppe Leser oder Nichtleser bzw. engagiertere oder distanziertere Anhänger dieses Zeitschriftentyps zusammenführt und diskutieren läßt.

Schließlich kann man auch Ergebnisse von Gruppen mit Lesern solchen mit Nichtlesern oder dies noch kombiniert mit anderen soziodemografischen Merkmalen analysieren bzw. als Pretest für eine Fragebogenbefragung nutzen.

Die wichtigste und allgemeinste Auswahlregel besagt, daß alle Personen in einer erklärten und bestimmaren Beziehung zum jeweiligen Thema stehen müssen.

Allerdings ist zu beachten, daß im Laufe der Diskussion mit „gleichgesinnten“ Teilnehmern nicht zu schnell ein Konsens gefunden wird. Konsens tötet die Diskussion und engt das Ergebnis in unzulässiger Weise ein. Ein ähnliches Phänomen ergibt sich auch bei Skalierung von eindeutigen Statements. Sie sind in Fragebogen eigentlich überflüssig. (Einheitliche Ergebnisse kommen allerdings bedauerlicherweise so manchem Auftraggeber bei der Untersuchung von Kommunikationsmaterial sehr entgegen!)

Die Entscheidung über die Zusammensetzung der Gruppe bestimmt auch indirekt die Zahl der durchzuführenden Gruppen.

Im allgemeinen werden Studien mit 2 bis 6 Gruppen durchgeführt. Diese Zahl reicht aus, um ein breites Antworten-Spektrum zu bekommen. Eine größere Anzahl von Gruppen – innerhalb eines Kultur- bzw. Sprachraums – ist aufgrund der entstehenden Kosten nicht mehr zu empfehlen.

Bei der Zusammensetzung der Gruppe sind in Übereinstimmung mit der Aufgabenstellung und der ihr zugrundeliegenden Zielgruppe

soziodemografische Merkmale:

- Geschlecht,
- Alter,
- soziale Schicht,
- Region etc.;

2) TAT = Thematischer Apperzeptions-Test.

psychografische Merkmale:

- bestimmte Einstellungen,
- Lifestyle

und

verhaltensrelevante Merkmale:

- Verwender/Nichtverwender/Kenner,
- zu berücksichtigen.

3.2 Die räumlichen Voraussetzungen

In den letzten Jahren haben sich die Anforderungen der Auftraggeber an die räumliche Ausstattung von Gruppendiskussionsräumen sehr gewandelt.

Gruppendiskussionsräume haben

- einen zusätzlichen Beobachtungsraum mit Einwegspiegel und Videoausrüstung,
- ausreichenden Platz für mehrere Beobachter,
- eine aufwendige Ton-/Bildausrüstung im eigentlichen Diskussionsraum mit Rückkopplungsmöglichkeit aus dem Beobachtungsraum,
- eine Übersetzerkabine,
- ein ausgeprägtes Catering-System etc.

Da diese Art von Ausstattung meist einen sehr elegant-kühlen und gleichzeitig labormäßigen Eindruck vermittelt, gehen einige Studios und Institute dazu über, neben diesem Standard gleichzeitig eine *familiär-private Raumatmosphäre* anzubieten.

Prinzipiell stellt sich hier die gleiche Frage wie bei Einzelinterviews:

- Sollten Gruppendiskussionen in der Wohnung der Zielperson, also in der privaten und damit gewohnten, vertrauten, den Lebensstil repräsentierenden Umgebung durchgeführt werden? Dies kann man beschreiben, fotografieren, protokollieren und so für die Auswertung nutzen (Nachteil: Es klingelt an der Tür, das Essen brennt an, die Kinder streiten sich und vor allem hört ein Partner die Fragen mit und gibt Kommentare.)

oder

- sollten Gruppendiskussionen in einem Befragungsstudio durchgeführt werden, wo es ein einheitliches, für jeden identisches Umfeld gibt?

Hier wirken sich dann in den Befragungsräumen meist Enge, Lichtverhältnisse, Improvisation bei der Vorlage des Untersuchungsmaterials etc. und gelegentlich auch die Durchführung mehrerer Befragungen in einem Raum in irgendeiner Weise auf das Antwortverhalten aus.

Gruppendiskussionsräume unterscheiden sich meist grundlegend von (Einzel-) Befragungsräumen im oben beschriebenen Sinne.

Eine Rückführung auf ein „Normalmaß“ ist vorläufig nicht zu erwarten, zumindest solange nicht endlich einmal jemand die Auswirkungen auf das Antwortverhalten und die Art der Antworten in unterschiedlich ausgestatteten Räumen untersucht und auch im Hinblick auf den Gruppendiskussionstourismus!

Die räumliche Ausstattung und die Zusatzforderungen an Bewirtung (nicht nur der Gesprächsteilnehmer) wirken sich natürlich auch auf die Kosten aus. Allein die Raummiete mit entsprechender technischer Ausstattung, Bewirtung und Einladung „normaler“ Gesprächsteilnehmer beträgt für eine zweistündige Gruppendiskussion etwa DM 2 000,-.

3.3 Zeitbedarf für Gruppendiskussionen

Im allgemeinen geht man davon aus, daß Gruppendiskussionen 1½ bis 2 Stunden Zeit in Anspruch nehmen.

„Kreativ-Gruppen“ können dabei auch als tagesfüllende Veranstaltung eingeplant werden, wobei jedoch ein deutlich erhöhter Kostenaufwand entsteht (ca. DM 10 000,- bis DM 12 000,-).

3.4 Datenschutz

Der Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V. (ADM) gibt zur Zeit gerade Richtlinien für die Durchführung von Gruppendiskussionen und vor allem die Weitergabe von Ton-/Bild-Aufzeichnungen, die dem Datenschutz und dem Standesrecht entsprechen, heraus. Sie sind in jedem Falle im übrigen auch von ausländischen Auftraggebern zu befolgen.

4 Fazit

Insgesamt sind also – ähnlich für Einzelinterviews und Explorationen – zahlreiche Einflußfaktoren auf das Antwort-Verhalten von Gruppendiskussionsteilnehmern zu berücksichtigen.

Die Gruppendiskussion hat methodisch ihren Sinn als „echter“ Pretest, der Sammlung von ersten psychologisch-soziologischen, soziografischen und verhaltensrelevanten Informationen über einen Untersuchungsgegenstand.

Das detaillierte Briefing vor der Gruppendiskussion ist wie bei jeder Methoden-Entwicklung von entscheidender Bedeutung.

Ein Fragebogen sollte erst nach Vorlage der Ergebnisse von Gruppendiskussionen entwickelt werden. Ideen für den Fragebogen und schwierige Formulierungen sollten in der Gruppendiskussion auf Verständlichkeit überprüft werden. Der Fragebogen ist dann als solcher einem Pretest im Feld zu unterziehen.

Kognitionspsychologische Ansätze zum Pretest von Fragebogen

Ein gesonderter Beitrag zu „kognitionspsychologischen Ansätzen zum Pretest von Fragebogen“ mag suggerieren, daß kognitionspsychologische Ansätze bei Pretests eher die Ausnahme als die Regel darstellen. Dies ist nicht so. Letztendlich beruhen alle der in diesem Band diskutierten Verfahren auf kognitionspsychologischen Ansätzen. Die Herausstellung der Kognitionspsychologie durch einen eigenen Beitrag entspringt vielmehr dem Anliegen, der derzeitigen Bedeutung kognitionspsychologischer Ansätze in der Umfragemethodologie Rechnung zu tragen und die Grundlagen dieses wissenschaftlichen Ansatzes darzustellen. Da die einzelnen Pretestverfahren Gegenstand eigenständiger Beiträge sind, werde ich mich in bezug auf einzelne Verfahren nur auf einige Anmerkungen zu Think-Aloud-Verfahren und Experimenten (Split-Ballot-Verfahren) beschränken. In diesem Rahmen werde ich zur Illustration einige ausgewählte Arbeiten der kognitionspsychologischen Forschung vorstellen. Obwohl sie alle eindeutig der Grundlagenforschung zuzurechnen sind, möchte ich sie in diesem Zusammenhang als Pretests verstehen. Im letzten Abschnitt dieses Beitrags wird diese Perspektive von Grundlagenforschung als Pretest explizit aufgegriffen.

1 Kognitionspsychologie und Umfragemethodologie

Die Umfragemethodologie war zwar implizit oder explizit schon immer kognitiv orientiert, aber die Fokussierung auf die kognitiven Prozesse ist eine relativ neue Entwicklung, deren Geburtsstunde vielleicht im Jahr 1984 liegt, als in Reaktion auf eine Konferenz „Cognitive aspects of survey methodology: Building a bridge between disciplines“ (Jabine, Straf, Tanur, Tourangeau 1984) das Interesse der Umfragemethodologen an kognitiven Prozessen – und umgekehrt das Interesse von Sozial- und Kognitionspsychologen an der Umfragemethodologie – explosionsartig zunahm. Vielleicht kann man die kognitive Wende in der Umfragemethodologie etwas überspitzt folgendermaßen darstellen. Frage und Antwort werden *nicht* mehr als bloße Reiz-Reaktionsverbindung aufgefaßt, was hieße, daß eine Frage eine Antwort bedingt, bzw. daß eine bestimmte Formulierung, Reihenfolge etc. eine bestimmte Antwortverzerrung nach sich zieht. Dagegen war die frühere Phase der Umfragemethodologie im wesentlichen am Phänomen solcher Antwortverzerrungen orientiert. Man identifizierte eine Fülle von Effekten, deren Erklärung aber relativ unsystematisch blieb. Kognitions- und Sozialpsychologen stellten in neuerer Zeit das Bild des Befragten als Informationsverarbeiter auf. Im Mittelpunkt des Interesses der kognitiv orientierten Umfragemethodologie stehen die kognitiven Prozesse, die beim Befragten in Reaktion auf eine Frage ablaufen und die Antwort bedingen.

Welches sind diese Prozesse? Zunächst müssen Befragte die Frage verstehen. Zum einen müssen sie sie semantisch verstehen, d. h., die Begriffe müssen für sie die gleiche Konnotation haben wie für den Forscher. Die Befragten müssen die Frage aber auch pragmatisch verstehen, d. h., sie müssen dekodieren, „was will der Interviewer eigentlich wissen?“. Hier zeigte sich die Forschungsperspektive fruchtbar, die Umfragesituation als natürliche

*) Dr. Michaela Wänke, Ruprecht-Karls-Universität Heidelberg.

Kommunikationssituation zu betrachten, in der entsprechende Konversationsnormen gelten. In natürlicher Kommunikation wird der Kommunikationserfolg unter anderem davon bestimmt, in welchem Ausmaß die Kommunikationspartner miteinander kooperieren, d. h. dem Partner das Optimum an Information mitteilen. Der englische Sprachphilosoph Grice (1975) identifizierte eine Reihe von Regeln, die diese Kooperation leiten. Zusammengefaßt läßt sich sagen, daß Beiträge so informativ wie möglich sein sollen, aber auch so wenig redundant wie möglich. Um beides zu erreichen, muß das Vorwissen des jeweiligen Partners berücksichtigt werden, denn es wäre redundant, dem Partner Information mitzuteilen, die dieser schon hat. Genausowenig würde ein Kommunikationspartner um Information bitten, über die er bereits verfügt. Auch in der nicht-natürlichen Situation der standardisierten Befragung können diese Konversationsnormen zum Tragen kommen. Strack, Schwarz und Wänke (1991) zeigten, daß sich die hohe Korrelation von berichtetem Lebensglück und berichteter Lebenszufriedenheit stark verminderte, wenn die Befragten beide Fragen in einem gemeinsamen Kontext wahrnahmen. Es scheint, daß die Befragten in diesem Fall stärker zwischen beiden Konzepten differenzierten, weil sie annahmen, daß nicht das gleiche gemeint sein konnte, wenn beide Fragen gemeinsam und unmittelbar aufeinanderfolgend präsentiert wurden. Zweimal hintereinander das gleiche zu fragen, würde den Konversationsnormen widersprechen. Der gleiche Prozeß, das Wahrnehmen eines Kommunikationskontexts und das dadurch ausgelöste Bestreben, Redundanz zu vermeiden, kann auch erklären, warum die Frage nach Teilaspekten (z. B. Ehezufriedenheit; Freigabe der Abtreibung bei medizinischer Indikation) vor der allgemeinen Frage (Lebenszufriedenheit; Freigabe der Abtreibung) zu Kontrasteffekten führen kann (z. B. Schwarz, Strack, Mai 1991; Schuman, Presser 1981). Man nimmt an, daß die Befragten bei den allgemeinen Fragen den Teilaspekt ausklammern, da sie sich dazu schon geäußert haben. Sie verstehen z. B. die Frage zur allgemeinen Lebenszufriedenheit, die auf eine Frage zur Ehezufriedenheit folgte als, „einmal abgesehen von Ihrer Ehezufriedenheit, die Sie gerade berichtet haben, wie zufrieden sind Sie mit ihrem Leben?“ (Schwarz, Strack, Mai 1991).

Nach der Interpretation der Frage müssen die Befragten im nächsten Schritt entweder eine Antwort, also ein Urteil abrufen oder eine Information, die es ihnen erlaubt, ein solches Urteil zu bilden. Dieses Urteil müssen sie dann in die vorgegebenen Antwortkategorien einpassen und gegebenenfalls editieren, wenn sie diese Antwort verschleiern wollen, z. B. aus Gründen der sozialen Erwünschtheit. (Zum Überblick der kognitiven Prozesse in der Befragung siehe Strack 1994.)

2 Leistungen von Pretestverfahren

Auf dieser kognitiven Perspektive, die salopp ausgedrückt werden könnte als „der Befragte denkt sich etwas bei der Beantwortung von Fragen“, basieren die meisten Pretestverfahren. Mit ihnen lassen sich die kognitiven sowie kommunikativen Prozesse der Befragung aufzeigen. So sollen diese Verfahren identifizieren, ob

- die Frage so verstanden wurde, wie vom Forscher beabsichtigt,
- Befragte überhaupt über Information verfügen, die sie benötigen, um die Frage zu beantworten,
- Befragte in der Lage sind, die benötigte Information abzurufen,

- Befragte in der Lage sind, die Information in die entsprechende Form zu übertragen, die nötig ist, um eine Antwort abzugeben,
- Befragte gewillt sind, ihr privates Urteil auch öffentlich preiszugeben.

2.1 Think-Aloud-Verfahren

Am deutlichsten wird diese kognitive Perspektive sicherlich im kognitiven Interview, auch Think-Aloud-Verfahren oder „verbal protocols“ genannt. Dabei werden die Befragten mehr oder weniger deutlich ermuntert, laut zu denken, und die Wege, die sie bei ihrer Antwort oder Urteilsfindung beschreiten, offen darzulegen. Ziel des Verfahrens ist es, etwas über die kognitiven Prozesse zu erfahren. So können Think-Aloud-Techniken zu klassischen Pretestzwecken eingesetzt werden, etwa um zu untersuchen, ob die Befragten tatsächlich die vom Frager intendierte Information abriefen, oder wie sie die Frage verstehen. Darüber hinaus findet diese Methode auch Anwendung in der Grundlagenforschung, um die der Fragebeantwortung zugrundeliegenden Prozesse aufzuspüren. Die Einzelheiten des Verfahrens sind in einem gesonderten Beitrag dargestellt. An dieser Stelle sind aber einige grundsätzliche Überlegungen zur Validität dieser Technik angebracht. Aus kognitionspsychologischer Sicht ist es zumindest problematisch zu unterstellen, daß Befragte introspektiv Zugang zu ihren kognitiven Prozessen haben und diese wiedergeben können.

Bei Fragen nach Verhalten können Think-Aloud-Techniken sehr gut funktionieren. Beispielsweise könnte ein solches Protokoll zur Frage, „wie oft im Monat essen Sie in einem Restaurant?“ folgende Schritte identifizieren: „Gestern war ich beim Italiener, und dienstags spiele ich Squash, da gehen wir immer hinterher etwas essen. Am Wochenende habe ich meine Tante zum Geburtstag in ihr Lieblingsrestaurant eingeladen. Das macht also dreimal. Aber das war eine untypische Woche, normalerweise gehe ich weniger oft aus, also vielleicht insgesamt so neun Restaurantbesuche im Monat.“ Beispiele dieser Art gaben wertvolle Hinweise über die Strategien, mit denen Befragte Verhaltenshäufigkeiten abriefen, z. B. unter welchen Bedingungen sie episodische Ereignisse abrufen, und unter welchen Bedingungen sie schätzen oder beide Strategien vermischen etc. (vgl. z. B. Blair, Burton 1987; Menon 1993; Bickart, Felcher 1996).

Problematischer ist der Fall bei Einstellungsfragen. Die Sozialpsychologie geht davon aus, daß Urteile meist in der Situation, in der sie gebraucht werden, also zum Beispiel in der Befragung, konstruiert werden. Dazu wird zu diesem Zeitpunkt verfügbare Information benutzt. Manche Information wird den Befragten chronisch verfügbar sein, d. h., sie wird dem Befragten zu diesem Thema immer einfallen, eine andere Information wird nur temporär verfügbar sein, weil sie durch den situativen Kontext, z. B. den Fragebogen, aktiviert wurde.

Zum Beispiel stimmten dem Recht auf freie Meinungsäußerung für alle Bürger deutlich weniger zu, wenn in einem Abstand von acht Fragen vorher im Fragebogen eine Frage zum Recht auf freie Meinungsäußerung des Ku-Klux-Klans versus der Elternvereinigung präsentiert wurde (Ottati, Riggle, Wyer, Schwarz, Kuklinski 1989). Offenkundig dachten die Befragten bei der Redefreiheit aller Bürger auch an den Ku-Klux-Klan, dem sie die Redefreiheit beschränken wollten. Diese Art von Assimilationseffekten finden sich allerdings nur, wenn die Befragten sich des Einflusses der Vorfrage nicht bewußt sind. Vermuten Befragte

potentielle Einflüsse, können sich die Effekte umdrehen (Strack, Schwarz, Bless, Kübler, Wänke 1993). So fanden auch Ottati u. a. nach der Frage zur Redefreiheit des Ku-Klux-Klans mehr Zustimmung zur allgemeinen Redefreiheit anstatt weniger, wenn die beiden Fragen nicht von acht Zwischenfragen getrennt wurden, sondern direkt aufeinander folgten. Eine Erklärung für diesen Einfluß des Abstands der Fragen läßt sich darin finden, daß im ersten Falle die entsprechenden Subgruppen zwar kognitiv aktiviert waren, den Befragten also leicht in den Sinn kamen, sie sich aber nicht bewußt waren, daß diese Beispiele ihnen nur deshalb in den Sinn kommen, weil sie in einer Vorfrage angesprochen wurden. Trennten keine Puffer die Einflußfrage von der Zielfrage, war dieser Einfluß aber offensichtlich, und die Befragten – bestrebt, ein unbeeinflusstes Urteil abzugeben – korrigierten ihr Urteil. Hätten im ersten Fall die Befragten also Wissen über den beeinflussenden Effekt der acht Fragen vorher stehenden Frage gehabt und äußern können, hätte sich der Assimilationseffekt gar nicht gezeigt, sondern ein Kontrasteffekt. Wir können annehmen, daß die Befragten keine Kognition über die beeinflussende Information hatten und sie auch nicht hätten äußern können.¹⁾

Individuen haben in der Regel extrem schlechten Zugang zu ihren Vorlieben, Verhaltenswahlen u.ä. und können deren Grundlagen introspektiv kaum erfassen (Nisbett, Wilson 1977). Sie haben auch keine Kognition darüber, daß ihr Urteil in guter Stimmung anders ausfällt als in schlechter Stimmung – ein Befund, der in der Sozialpsychologie oft und bei unterschiedlichen Urteilsbereichen gezeigt wurde (zum Überblick: Clore, Schwarz, Conway 1994). Sie wissen auch nicht, wie andere die Informationsverarbeitung begleitende Erfahrungen, wie zum Beispiel die Leichtigkeit, mit der relevante Information in den Sinn kommt, das Urteil beeinflussen. Dabei zeigen sich auch hier eklatante Effekte.

In einer Untersuchung baten wir (Wänke, Bless, Biller 1995) eine Hälfte von Versuchspersonen, Gründe zu nennen, die für eine Benutzung des öffentlichen Personennahverkehrs (ÖPNV) sprechen. Die andere Hälfte baten wir, Gründe zu nennen, die dagegen sprechen. Da nun für die eine Gruppe Pro-Argumente aktiviert waren und für die andere Gruppe Contra-Argumente, könnte man erwarten, daß dementsprechend die experimentell manipulierte Pro-Gruppe eine positivere Einstellung zur Benutzung des ÖPNV berichtete als die experimentell manipulierte Contra-Gruppe. Dieser Einstellungsunterschied würde den Einfluß der verfügbaren Information auf das Einstellungsurteil widerspiegeln. Zusätzlich aber schafften wir experimentelle Bedingungen, die es jeweils einer Hälfte der Pro- sowie der Contra-Gruppe sehr leicht machte, Gründe zu nennen. Den jeweils anderen Hälften machten wir es schwer. Dabei zeigte sich, daß der Effekt, Pro-Argumente führen zu positiver Einstellung und Contra-Argumente zu negativer Einstellung, nur auftrat, wenn es den Befragten leicht fiel, Gründe zu finden. Die Befragten, denen es experimentell erschwert wurde, Pro-Argumente zu finden, berichteten eine wesentlich negativere Einstellung als die Befragten, denen es experimentell erleichtert wurde. Umgekehrt berichteten die Befragten, denen es experimentell erschwert wurde, Contra-Argumente zu nennen, eine wesentlich positivere Einstellung als die Befragten, denen es experimentell erleichtert wurde. Dabei muß betont

1) Es läßt sich einwenden, daß Befragte gar kein Wissen über den beeinflussenden Effekt haben müssen, um diesen im kognitiven Interview zu entdecken. Es würde ausreichen festzustellen, daß Befragte unter anderem an den Ku-Klux-Klan dachten und deshalb zu negativen Urteilen kamen. Allerdings läßt sich nicht ausschließen, daß durch die Verbalisierung die Inhalte auf eine bewußtere Ebene gebracht werden und deshalb anders verarbeitet werden als in der normalen Situation.

werden, daß die Befragten, denen die Generierung der Argumente erschwert wurde, trotzdem mehr Argumente nannten als die Befragten, denen es erleichtert wurde. Der Einfluß der subjektiv erlebten Leichtigkeit war also nicht durch die generierte Anzahl vermittelt. Eher scheint die Leichtigkeit, mit der bestimmte Inhalte in den Sinn kamen, die Wirkung dieser Inhalte auf das Urteil zu moderieren. Es ist zu bezweifeln, daß die Befragten sich dieser vermittelnden Rolle der subjektiv erlebten Leichtigkeit bewußt waren.

Die meisten der bei der Urteilsbildung involvierten Prozesse sind introspektiv nicht zugänglich. Schlimmer noch, in einem eindrucksvollen Forschungsprogramm zeigten Wilson und Kollegen, daß die Aufforderung, Einstellungsurteile zu begründen oder zu erklären, nachfolgendes Verhalten und damit in Verbindung stehende Einstellungsurteile sehr stark beeinflusste (Wilson, Dunn, Kraft, Lisle 1989; Wilson, Hodges 1992). Die Bearbeitung einer Frage im Think-Aloud-Interview kann also möglicherweise Verzerrungen nachfolgender Antworten nach sich ziehen, die sich in der realen Befragung nicht ergeben würden. Die Validität des Verfahrens ist also nicht unbedingt gegeben.

2.2 Das Experiment

Was vielleicht bis hierher durch die geschilderten Experimente schon angeklungen ist und im folgenden noch deutlich werden wird, ist, daß Einflußgrößen, also Variablen, die die Antwort beeinflussen, am besten und methodisch am saubersten identifiziert werden können, indem man die vermeintliche Einflußgröße variiert und den Effekt der Variation auf die Antwort mißt. Befürchtet man beispielsweise, daß in einem Fragebogen die Frage 23 eventuell die Antworten zu Frage 24 beeinflusst, dann gibt es keine bessere Alternative, als die Fragereihenfolge zu variieren und so den eventuellen Reihenfolgeeffekt zu überprüfen. Liegen in der Tat Reihenfolgeeffekte vor, und man überlegt sich, ob der Einschub einiger Pufferfragen zwischen den beiden Fragen diesen Einfluß ausmerzen könnte, dann muß diese Hypothese getestet werden, indem zusätzlich zu den zwei Reihenfolgebedingungen eine Bedingung mit Pufferfragen gegeneinander getestet werden.

3 Grundlagenforschung als Pretest

Ein systematisches Testen vermeintlicher Einflußgrößen setzt allerdings Hypothesen über deren eventuelle Effekte voraus. Diese Hypothesen können begründet sein in der Erfahrung, den Hinweisen durch Interviewer oder durch die Kenntnis kognitionspsychologischer Theorien, Modelle und Befunde. In jedem Fall aber ist der Test theoriengeleitet. Worauf ich hinaus will, ist, daß es unabhängig von den Pretestinstrumenten zwei Vorgehensweisen gibt, die sich ergänzen und sich im Interesse der Qualität von Umfragedaten ergänzen müssen. Zum einen Pretests, die eine Art „final check“ darstellen. In diesen Fällen wird ein im Prinzip fertiger Fragebogen – erstellt nach allem Expertenwissen – unter Realbedingungen²⁾ verwendet und offensichtliche Schwierigkeiten werden so identifiziert. Mittels Verhaltenskodierung läßt sich beispielsweise feststellen, bei welchen Fragen Probleme auftauchen – etwa Verständnisprobleme beim Befragten, Abweichungen des Interviewers vom Standardtext, ob die Befragten die Frage so verstehen wie intendiert, ob die Antworten so detailliert wie

2) Beim kognitiven Interview sind es keine Realbedingungen.

gewünscht ausfallen etc. (z. B. Fowler, Cannell 1996). Es läßt sich aber nicht feststellen, ob Reihenfolgeeffekte auftreten oder wie die vorgegebene Antwortskala die Antworten beeinflußt etc. Auch folgender Effekt ließe sich mittels solcher Pretests nicht identifizieren.

Wir fragten ca. 200 Befragte nach einem Vergleich der Umweltverschmutzung durch Industrie und durch Verkehr (vgl. Wänke, Schwarz, Noelle-Neumann 1995). Die Befragten wurden in einer Bedingung gefragt: „Würden Sie sagen, der Verkehr trägt eher mehr oder eher weniger als die Industrie zur Umweltverschmutzung bei?“³⁾ 45 % der Befragten sahen in dem Verkehr den größeren Umweltverschmutzer gegenüber 32 %, die eher der Industrie einen größeren Verursacheranteil zusprachen. In einer anderen Bedingung aber variierten wir die Vergleichsrichtung und fragten: „Würden Sie sagen, die Industrie trägt eher mehr (weniger) oder eher weniger (mehr) als der Verkehr zur Umweltverschmutzung bei? Auf diese Frage hin wurde eher die Industrie von 57 % als der größere Verursacher von Umweltverschmutzung beurteilt und nur 24 % sahen im Verkehr den Hauptverursacher. Die in der Frage vorgegebene Vergleichsrichtung bewirkte demnach eine Umkehrung der Rangordnung. Ähnliche Effekte der Vergleichsrichtung, ob A mit B oder B mit A verglichen wird, wurden mehrfach in unterschiedlichen Befragungsmodi, Inhaltsgebieten und Antwortalternativen repliziert. Dieses Ergebnis hat uns nicht überrascht. Genau diese Unterschiede und diese Richtung der Abweichung hatten wir erwartet aufgrund der kognitiven Prozesse, die bei Vergleichen ablaufen (vgl. Wänke, Schwarz, Noelle-Neumann 1995; Wänke 1993). Die psychologischen Modelle über kognitive Prozesse sagen solche Asymmetrien vorher, und unsere Experimente testeten diese Vorhersagen. Diese Art von Tests, ein theoriengeleitetes Testen von Hypothesen, möchte ich gegenüber dem, was bisher unter Pretest verstanden wird, als den „anderen Pretest“ bezeichnen.

Liefert uns die Grundlagenforschung Pretestdaten? In gewisser Weise tut sie das, denn was des einen Pretest ist, ist des anderen Hauptuntersuchung. Die wenigen Beispiele machten hoffentlich deutlich, daß wir beide Arten von Pretests brauchen, die Grundlagenforschung und die spezifische Überprüfung am eigentlichen Fragebogen. Die beiden Arten betreffen allerdings unterschiedliche Zeitstufen in der Planung und unterschiedliche Abstraktionsstufen. Es wird kein Weg daran vorbeiführen zu überprüfen, ob Fragen verständlich und eindeutig sind, ob der Ablauf des Fragebogens unproblematisch ist, ob der Befragte überhaupt über Informationen verfügt, die Frage zu beantworten etc. Von entscheidender Bedeutung ist dabei, daß Fragen immer in dem Kontext getestet werden, in dem sie später präsentiert werden. Ein Pretest ist nur so gut, wie die Bedingungen seiner Durchführung den realen Bedingungen bei der tatsächlichen Hauptuntersuchung entsprechen. Dieser „final check“ wird daher in einer späten Phase und konkret am jeweiligen Fragebogen erfolgen müssen.

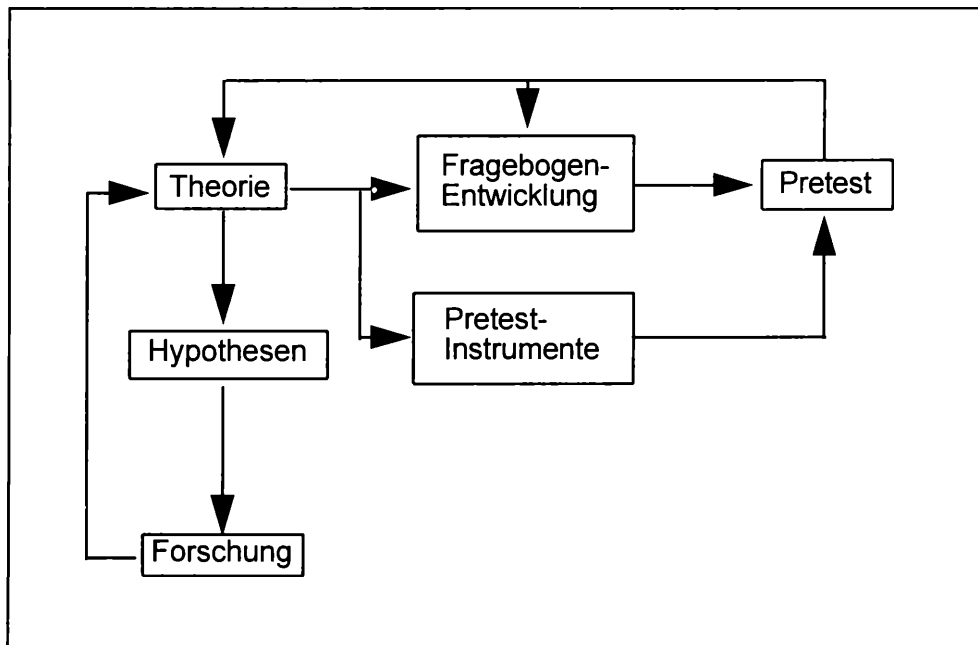
Vorher aber gilt es zu überlegen, mit welchen Effekten zu rechnen ist, welche Frageform, welche Antwortalternative, welche Frageformulierung, welche Fragereihenfolge etc. zu wählen ist, um zu – im Sinne des Forschungsinteresses – möglichst validen Daten zu kommen. Für diese Zwecke muß auf die „anderen Pretests“ zurückgegriffen werden. Keinesfalls heißt das aber, daß zu jeder Studie spezifische Untersuchungen anfallen müssen. Sicherlich wäre damit die Kapazität jedes Instituts überfordert. Die Grundlagenforschung, sei es im

3) Innerhalb dieser Bedingung wurde auch noch die Reihenfolge der Antwortalternativen variiert, so daß die Hälfte gefragt wurde: „Würden Sie sagen, der Verkehr trägt eher weniger oder eher mehr als die Industrie zur Umweltverschmutzung bei?“ Diese Reihenfolgemanipulation zeigte jedoch keinen Effekt.

Laborexperiment oder in Umfragen unter Realbedingung, liefert ständig neue Erkenntnisse, die sich die angewandte Forschung nutzbar machen kann. In diesem Sinne liefert Grundlagenforschung *unspezifische Pretests*. Unspezifisch deshalb, weil diese Untersuchungen nicht an einer spezifischen Frage oder einem spezifischen Fragebogen orientiert sind, sondern statt dessen die allgemeinen in der Befragung ablaufenden Prozesse untersuchen. Vielleicht werden publizierte Forschungsergebnisse nicht in allen Fällen übertragbar sein. In solchen Fällen stehen dann weitere in der Tat spezifische Pretests an. Aus diesen gewinnen wir aber hoffentlich nicht nur die Erkenntnis, ob z. B. die Frage 23 vor der Frage 24 präsentiert werden muß, sondern auch generell ein besseres Verständnis von – wie in diesem Beispiel – Reihenfolgeeffekten. Ergebnisse, die dem Praktiker etwas über die Frage 23 sagen, sagen der Umfragemethodologie vielleicht etwas über allgemeinere Prozesse.

In diesem Sinne gilt mein Plädoyer einer stärkeren Kooperation der Grundlagenforschung und der angewandten Forschung und Praxis. Das heißt zum einen eine stärkere Einbeziehung und Nutzung der Grundlagenforschung, zum anderen wäre es wünschenswert, wenn auch die Praktiker die methodologischen Implikationen ihrer Befunde stärker in allgemeine Modelle der Frage-Antwort-Prozesse einbrächten. Der Nutzen für die Umfragemethodologie liegt auf der Hand. Die folgende Abbildung 1 zeigt, wie sich die „beiden Arten von Pretests“ gegenseitig ergänzen.

Abbildung 1
Der Zusammenhang von Grundlagenforschung
und angewandter Forschung in der Umfragemethodologie



Allgemeine Phänomene und Prozesse können nur durch theoriengeleitetes Vorgehen entdeckt und erforscht werden. Mit deren Verständnis können wir aber auch bessere Fragen entwickeln. Gleichmaßen lassen sich aus den theoretischen Erkenntnissen verbesserte Pretestinstrumente entwickeln, die eventuell auftretende Probleme identifizieren können. Umgekehrt verweisen solche Pretests wieder auf Phänomene, die systematisch untersucht werden müssen. Damit schließt sich der Kreislauf, ein – im doppelten Wortsinn – Kreislauf der Wissen(-)schaff(f)t.

Literaturhinweise

Bickart, B., Felcher, M. (1996): Expanding and enhancing the use of verbal protocols in survey research, in: N. Schwarz, S. Sudman (Hrsg.): Answering questions: Methodology for determining cognitive and communicative processes in survey research, San Francisco.

Blair, E., Burton, S. (1987): Cognitive processes used by survey respondents to answer behavioral frequency questions, in: Journal of Consumer Research, 14, S. 280 – 288.

Clore, G., Schwarz, N., Conway, M. (1994): Cognitive causes and consequences of emotion, in: R. S. Wyer, T. K. Srull (Hrsg.): Handbook of Social Cognition (2. Auflage, S. 323 – 417), Hillsdale, N.J.

Fowler, F., Cannell, C. (1996): Using behavioral coding to identify cognitive problems with survey questions, in: N. Schwarz, S. Sudman (Hrsg.): Answering questions: Methodology for determining cognitive and communicative processes in survey research, San Francisco.

Grice, H. P. (1975): Logic and Conversation, in: P. Cole, J. L. Morgan (Hrsg.): Syntax and Semantics, Band 3: Speech Acts, New York, S. 41 – 58.

Jabine, B., Straf, L., Tanur, J., Tourangeau, R. (1984): Cognitive aspects of survey methodology: Building a bridge between disciplines, Washington, D.C.

Menon, G. (1993): The effects of accessibility of information in memory on judgments of behavioral frequencies, in: Journal of Consumer Research, 20, S. 431 – 440.

Nisbett, R., Wilson, T. (1977): Telling more than we can know: Verbal reports on mental processes, in: Psychological Review, 84, S. 231 – 259.

Ottati, V. C., Riggle, E. J., Wyer, R. S., Schwarz, N., Kuklinski, J. (1989): The Cognitive and Affective Bases of Opinion Survey Responses, in: Journal of Personality and Social Psychology, 57, S. 404 – 415.

Schuman, H., Presser, S. (1981): Questions and Answers in Attitude Surveys, New York.

Schwarz, N., Strack, F., Mai, H.-P. (1991): Assimilation and Contrast Effects in Part-Whole Question Sequences: A Conversational Logic Analysis, in: Public Opinion Quarterly, 55, S. 3 – 23.

Strack, F. (1994): Zur Psychologie der standardisierten Befragung, Berlin.

Strack, F., Schwarz, N., Bless, H., Kübler, A., Wänke, M. (1993): Awareness of the influence as a determinant of assimilation vs. contrast, in: *European Journal of Social Psychology*, 23, S. 53 – 62.

Strack, F., Schwarz, N., Wänke, M. (1991): Semantic and pragmatic aspects of context effects in social and psychological research, in: *Social Cognition*, 9, S. 111 – 125.

Wänke, M. (1993): Vergleichsprozesse bei evaluativen Urteilen: Der Einfluß der in der Frage vorgegebenen Vergleichsrichtung, New York.

Wänke, M., Bless, H., Biller, B. (1995): Subjective experience versus content of information in the construction of attitude judgments, unveröffentlichtes Manuskript, Universität Heidelberg.

Wänke, M., Schwarz, N., Noelle-Neumann, E. (1995): Question wording in comparative judgments: Understanding and manipulating the dynamics of the direction of comparison, in: *Public Opinion Quarterly*, S. 347 – 372.

Wilson, T. D., Dunn, D. S., Kraft, D., Lisle, D. J. (1989): Introspection, attitude change, and attitude-behavior consistency: The disruptive effects of explaining why we feel the way we do, in L. Berkowitz (Hrsg.): *Advances in Experimental Social Psychology*, Bd. 22, Orlando, FL, S. 287 – 343.

Wilson, T. D., Hodges, S. D. (1992): Attitudes as temporary constructions, in: L. Martin, A. Tesser (Hrsg.): *The construction of social judgment*, Hillsdale, N J., S. 37 – 65.

Die Fragebogenkonferenz

Ich möchte eine persönliche Bemerkung vorausschicken: Ich war sehr überrascht, als ich Anfang des Jahres erfuhr, daß im Statistischen Bundesamt eine wissenschaftliche Arbeitstagung zum Thema Fragebogenentwicklung und Pretest angesetzt war. Ich hatte in den letzten Jahrzehnten zunehmend die Hoffnung aufgegeben, daß sich irgendwo ein ernsthaftes wissenschaftliches Interesse am Fragebogen entwickelt.

Jahrzehntlang hatte ich es hinausgeschoben, mein 1963 erschienenes Buch „Umfragen in der Massengesellschaft. Einführung in die Methoden der Demoskopie“ (Noelle 1963/1976) zu überarbeiten. 1976, als die 7. Auflage in Rowohlts Deutscher Enzyklopädie erschien, und es schon Übersetzungen in vier Sprachen gab, wechselte ich von Rowohlt zum Deutschen Taschenbuch-Verlag und schloß 1977 einen Vertrag über eine aktualisierte, stark überarbeitete Fassung. Ich rechnete mit großen methodischen Fortschritten – nicht nur technischen, sondern vor allem im Fragebogen, aber es kam nichts. Zwanzig Jahre nach Stanley L. Paynes (1951) „The Art of Asking Questions“ – Die Kunst, Fragen zu stellen –, 25 Jahre nach Hadley Cantrils (1947) „Gauging Public Opinion“ aus der Mitte der vierziger Jahre, wo ich die Methode des Split-Ballot kennenlernte, veröffentlichte ich 1970 im Public Opinion Quarterly einen Aufsatz über Ergebnisse gegabelter Befragung zur Fragebogenmethodologie: „Wanted: Rules For Wording Structured Questionnaires“ (Noelle-Neumann 1970).

Ich nahm mir vor, mit amerikanischen Kollegen zusammen ein umfassendes Standardbuch zur Fragebogenmethodologie zu schreiben, das in Nordamerika und in Europa brauchbar sein sollte. Mitte der siebziger Jahre lud ich Howard Schuman, Professor für Soziologie am Institute for Social Research and Department of Sociology, University of Michigan, und seit 1987 Editor des Public Opinion Quarterly, der damals ein Forschungsfreijahr in England verbrachte, nach Allensbach ein und schlug ihm in einem Abendgespräch solch ein gemeinsames Buch vor. Er lehnte ab und begründete das: „Sie sind uns zu weit voraus.“ Für mich war eine solche Begründung unbegreiflich, ich war sehr entmutigt.

1977 schloß ich auch mit der University of Chicago Press einen Vertrag für die englische Übersetzung des Methodenbuches von 1963, aber ich zögerte die Arbeit hinaus – ebenso wie beim Deutschen Taschenbuch-Verlag, ich dachte, es müßte auf breiter Front ein Interesse für Fragebogen-Methodologie einsetzen.

Längst hatten wir in Allensbach am Beispiel der Mediaforschung nachgewiesen, daß weder Stichprobenmethode oder Prinzipien der Interviewerorganisation noch irgendein anderer Faktor die Ergebnisse der Umfrageforschung so sehr wie der Fragebogen beeinflusst.

Norman Bradburn, Seymour Sudman, Stanley Presser und Howard Schuman begannen schließlich in den siebziger Jahren, bestimmte wichtige, aber doch sehr kleine Segmente der Fragebogenmethode zu untersuchen (Sudman, Bradburn 1974; Bradburn, Sudman u. a. 1979; Sudman, Bradburn 1982; Schuman, Presser 1981).

*) Prof. Dr. Dr. h.c. Elisabeth Noelle-Neumann, Institut für Demoskopie, Allensbach.

Norman Bradburn, Direktor des National Opinion Research Center (NORC) der University of Chicago, erkannte das große Potential des „Huckepack-Verfahrens“ für die Methodenlehre: die Nutzung von Repräsentativumfragen der Auftragsforschung zum Einbau von Methodenexperimenten in den Fragebogen. Aber es erwies sich als unmöglich, ein breites Forschungsprogramm von Split-Ballot-Experimenten, wie ich es 1974 auf dem Jahreskongreß der American Association of Public Opinion Research (AAPOR) in Lake George/USA vorgetragen hatte (Noelle-Neumann 1974)¹⁾, von der National Science Foundation als Forschungsprojekt bewilligt zu bekommen. Norman Bradburn erklärte außerdem, daß im NORC vom technischen Personal enormer Widerstand gegen Split-Ballot-Umfragen geleistet werde.

Als weitere fünfzehn Jahre vergangen waren, Anfang der neunziger Jahre, war mir klar, daß ich mit der Bearbeitung des Methodenbandes nicht länger warten dürfte. Ich benutzte meine Eric Voegelin-Gastprofessur in München 1993/94, um wieder eine ganzsemestrige Methodenvorlesung zu halten als Vorbereitung für die neue Buchfassung. Ich lud Thomas Petersen, der mir schon bei den Münchner Vorlesungen assistiert hatte, ein, als Koautor für die Neubearbeitung des Methodenbandes mitzuwirken, und ich versprach dem Deutschen Taschenbuch-Verlag, das Manuskript im Herbst 1995 rechtzeitig abzuliefern für ein Erscheinen im Frühjahr 1996.

Da erreichte mich die Nachricht von der wissenschaftlichen Tagung im Statistischen Bundesamt zu Problemen des Fragebogens und des Pretests.

Ich meldete sofort bei Peter Mohler, Zentrum für Umfragen, Methoden und Analysen (ZUMA), meinen Wunsch nach Beteiligung an: Nicht, weil ich glaubte, daß ich einen wichtigen Beitrag leisten könnte, sondern vor allem, um mit den an einem solchen Thema interessierten Kollegen zusammenzutreffen und ihre Referate zu hören, bevor das neue Manuskript zum Methodenbuch abgeschlossen sein würde.

Natürlich will ich Ihnen auch einiges berichten, aber mehr in der Art von Notizen. Hoffentlich wird vom Institut für Forschung, Aus- und Fortbildungsaufgaben in der Bundesstatistik des Statistischen Bundesamtes in einem Jahr abermals eine solche wissenschaftliche Tagung zu Fragebogenproblemen angesetzt. Davon würde eine außerordentlich anregende Wirkung ausgehen – und in einem Jahr wären wir schon viel weiter. Vielleicht steht dann schon die Gliederung des geplanten Standardwerkes zur Fragebogenmethode fest, bei dem wir mit Norbert Schwarz, University of Michigan, Institute of Social Research, und anderen Kollegen zusammenarbeiten wollen.

Wie notwendig ein solches Buch sein würde, wurde mir vor einigen Jahren schockartig klar: Damals veranstaltete die „Wirtschaftswoche“ eine Telefonumfrage bei Umfrage-Instituten und erkundigte sich auch beim Allensbacher Institut, wieviel des Budgets eines Auftrages für Markt- oder Meinungsforschung durchschnittlich für die Fragebogenentwicklung aufgewendet werde.

1) Dieser Vortrag ist nicht veröffentlicht, der vollständige Text und die Tabellen sind aber im Allensbacher Archiv greifbar: „Draft on an International Research Program with Regard to Questionnaire Methodology“ (31. Mai 1974).

Ich nannte den Betrag, der bei den regelmäßigen Allensbacher betriebswirtschaftlichen Nachkalkulationen festgestellt wird: etwa 25 % des Budgets.

Als einige Wochen später die „Wirtschaftswoche“ die Ergebnisse ihrer Telefonumfrage in einer Tabelle veröffentlichte, las ich:

Institut A:	2 %,
Institut B:	1 %,
Institut C:	3 %,
Institut für Demoskopie Allensbach:	25 %,
Institut E:	2 %

und so fort.

Wenn vom Budget 1 oder 2 % in die Fragebogenentwicklung gehen, dann zeigt das natürlich nur an, daß der Fragebogen völlig vernachlässigt wird. Unter dem Druck der weit verbreiteten Kriterien für Auftragsvergabe in der Markt- und Meinungsforschung – billiger, schneller – und der Unmöglichkeit, die Kosten für Stichprobe, Interviewer, Aufbereitung und Computer-Tabellenbände herabzudrücken, werden die Kosten da eingespart, wo es praktisch niemand bemerkt, nämlich am Fragebogen. Oft gibt es noch dazu das großzügige Angebot, der Auftraggeber dürfe sogar den Fragebogen selbst schreiben.

Die Zusammenstellung der Programmfragen, der Fragen, die durch die Untersuchung bearbeitet werden sollen, ist in all solchen Fällen bereits der Fragebogen. Eine Unterscheidung zwischen Programmfragen und Testfragen, die Übersetzung der Programmfragen in Testfragen, findet nicht statt. Der Auftraggeber ist beruhigt. Er findet seine Programmfragen im Fragebogen und sogar genau in den Formulierungen und mit der Verwendung der Fachbegriffe, die ihm als Experten richtig erscheinen.

In vielen Instituten entwickelt sich vor allem eine hohe Sensibilität, Fragen zu vermeiden, an denen sich die Validität der Ergebnisse kontrollieren ließe. Im übrigen schreibt ein Mitarbeiter, der Projektleiter, den Fragebogen. Pretests sind nicht erforderlich, um es in den Worten von Herrn Hans Akkerboom, Centraal Bureau voor de Statistiek, Heerlen, Niederlande, auszudrücken: „Der Fragebogen: Das ist eine Sache für Samstagnachmittag.“

Aber wie entsteht nun umgekehrt ein 25prozentiger Budgetanteil? Soviel aufzuwenden, das ist das Ergebnis der Entdeckung, daß die Qualität der Untersuchung von nichts mehr abhängt als von der Qualität des Fragebogens. Aus dieser Einsicht erwachsen zahlreiche kosten-treibende Elemente.

Teuer ist vor allem die Fragebogenkonferenz. Die Einrichtung der Fragebogenkonferenz ist ursprünglich durch das Modell der Redaktionskonferenz im Journalismus angeregt worden. Sie besteht aus drei oder vier auf Fragebogenentwurf spezialisierte Mitarbeiter und ist verantwortlich für *jeden* Fragebogen, der im Institut entsteht, unabhängig von der Art und Größe des Forschungsprojekts, unabhängig davon, wer der Projektleiter im Institut ist.

Fragebogenkonferenzen gibt es bis heute praktisch in keinem der großen Umfrageinstitute der Welt, sie sind bisher noch eine Spezialität des Allensbacher Instituts. Es ist darum wichtig, die Gründe zu nennen, warum es im Interesse der Qualität der Umfrageforschung grund-

sätzlich Fragebogenkonferenzen geben sollte – so grundsätzlich wie eine statistische Abteilung oder eine Interviewer-Organisation oder Computer zur Datenerfassung und -verarbeitung.

Die wichtigsten Gründe sind:

1. Nietzsche hat einmal geschrieben: „Alles Sehen ist perspektivisch“.

Die Fragebogenkonferenz wird so zusammengesetzt, daß möglichst verschiedene Perspektiven in ihr vertreten sind: Männer und Frauen, Menschen ganz verschiedenen Alters, verschiedene Studienfächer – ein abgeschlossenes Studium ist in der Regel die Voraussetzung für alle Mitglieder der Fragebogenkonferenz. Wünschenswert ist, daß die Mitglieder der Fragebogenkonferenz aus verschiedenen Regionen Deutschlands stammen, aus verschiedenen sozialen Schichten. Nicht nur wünschenswert, sondern notwendig ist, daß sie sich in ihrer politischen Orientierung unterscheiden. Das ist nahezu die einzige Möglichkeit, politisch einseitige Fragebogen zu vermeiden. Die Erfahrung zeigt, daß eine einseitige Frageformulierung als neutral empfunden wird, wenn sie in der Richtung der eigenen Überzeugungen verzerrt ist. Dagegen wird die Einseitigkeit sofort wahrgenommen, wenn die Formulierung eigenen Überzeugungen zuwiderläuft. Durch eine Fragebogenkonferenz, deren Mitglieder verschiedene politische Überzeugungen haben, ist gesichert, daß Einseitigkeit bemerkt wird, mal von der einen, mal von der anderen Seite, und damit korrigiert werden kann.

Durch die verschiedenen Perspektiven, Lebenserfahrungen, durch verschiedenes Fachwissen der Mitglieder der Fragebogenkonferenz wird außerdem eine größere Zahl von Aspekten, die für die Untersuchungsaufgabe wichtig sein können, eingebracht; es leuchtet sicher unmittelbar ein, daß drei oder vier Personen mit sehr verschiedenem Hintergrund mehr einfällt, was den Fragebogen anreichert, als einer Person, die, wie bisher noch üblich, allein den Fragebogen entwirft.

Einiges müssen die Mitglieder der Fragebogenkonferenz gemeinsam haben: sprachliche, journalistische, literarische Fähigkeiten, gute Allgemeinbildung, gute aktuelle Information im Bereich von Politik, Wirtschaft, Kultur, und zwar nicht nur aus dem Fernsehen und Radio, sondern aus Tages- und Wochenzeitungen und Monatszeitschriften. Schließlich ist erforderlich die Fähigkeit zu Teamwork, Selbstkritik, Disziplin, Toleranz.

2. Fragebogenentwicklung verlangt Spezialkenntnisse, wie man sie nur durch dauernde Beschäftigung mit Fragebogen – Entwurf und Testen – gewinnt.

Die Kunst der Übersetzung von Programmfragen – das sind die Untersuchungsaufgaben – in Testfragen, wie sie im Fragebogen selbst stehen, verlangt auch Kenntnis der zahlreichen Fragemodelle, die entwickelt wurden. Wahrscheinlich käme man leicht auf 50 verschiedene Modelle. Der Fragebogenspezialist muß wissen, unter welchen Umständen welche Modelle zu verwenden sind.

Erforderlich ist außerdem die Kenntnis der Interviewer-Situation, in der die Fragen sich bewähren sollen, so daß der Interviewer sich tatsächlich an seine Vorschrift hält: „Fragen wörtlich vorlesen, bei Rückfragen die Frage einfach zum zweiten Mal vorlesen, keine eigenen Erklärungen“.

Erforderlich ist auch die Kenntnis von Daten-Aufbereitung und -analyse. Beim Aufstellen des Fragebogens müssen die technischen und geistigen Erfordernisse der späteren Verarbeitungsstufe vor Augen stehen.

Notwendig sind auch Spezialkenntnisse über den besten Aufbau eines Fragebogens, Eisbrecherfragen, Pufferfragen, Spielfragen zur Belebung, zur Motivation der Befragten. Kenntnis, welche Fragen am Anfang, welche besser am Ende des Interviews stehen.

Dazu kommt die Beherrschung der Fragebogentechnik, Filterführung, Faktfragen, die im Wortlaut zu stellen sind, informelle, an keinen bestimmten Wortlaut gebundene Fragen, Interviewer-Beobachtungen, ohne dazu eine Frage zu stellen, kurze Interviewer-Instruktionen.

Zur Fragebogentechnik, die man beherrschen muß, gehört auch die Beachtung von Kontexteffekten: Wie weit Fragen, die vor – oder sogar nach! – Kernfragen gestellt werden, die Ergebnisse der Kernfragen beeinflussen können. Ausstrahlungseffekte von Fragen auf nachher (oder sogar vorher!) gestellte Fragen gehören zu den größten Fehlerquellen speziell bei Trendanalysen, wenn nämlich Trendfragen wörtlich in späteren Umfragen wiederholt werden, um einen Trend zu beobachten, aber das Frage-Umfeld und damit die Assoziationen, der Bezugsrahmen, auf den hin die Frage beantwortet wird, wechselt.

Zur Fragebogentechnik gehört auch das Wissen von der Anlage kontrollierter Feldexperimente – „Split-Ballot“, „gegabelte Befragung“ – im Fragebogen, um die Auswirkung von Reihenfolge-Effekten, von Alternativen, Listen, Umfeldeffekten, einzelnen Worten oder Formulierungen der Frage zu kontrollieren.

Zur Fragebogentechnik gehört schließlich auch die Kenntnis der Modelle zur Aktivierung des Befragten, bei denen nicht der Interviewer sortiert, schreibt, prüft, sondern, wo immer möglich, der Befragte selbst aktiv werden kann. Der Befragte soll beschäftigt werden, er soll selbst etwas tun.

Alle Mitglieder der Fragebogenkonferenz sind vertraut mit diesen Aufgaben, aber erst in der Zusammenarbeit werden die optimalen Lösungen gefunden.

3. Nicht alle Fragen, die durch eine Umfrage geklärt werden sollen, müssen übersetzt werden. Es gibt Fakt-Fragen, die ohne weiteres direkt gestellt werden können: „Haben Sie einen Führerschein?“, gegebenenfalls: „Welche Klasse?“

Viele Fragen gehören zum Typ der „Abstimmungsfrage“, wie man sie auch in einer Abstimmung der ganzen Bevölkerung vorlegen könnte: „Ja/Nein“, „Dafür/Dagegen“, „Gute Meinung/Keine gute Meinung“. Abstimmungsfragen mit ihren Ergebnissen sind es in erster Linie, die heute die Medien überschwemmen, oft mit bemerkenswert verschiedenen Ergebnissen. Um nur ein Beispiel zu geben: EMNID fragte für den Fernsehsender n-tv im Juli 1995: „Sind Sie für den ‚Großen Lauschangriff‘, also für das Abhören in Wohnungen, oder sind Sie dagegen?“ Ergebnis für Gesamtdeutschland: „Dafür“ 35 %, „Dagegen“ 63 %, „Keine Angabe“ 2 %.

Allensbach fragte für die FAZ: „Sind Sie dafür, daß das Abhören in Privatwohnungen erlaubt wird, damit man das organisierte Verbrechen wirksamer bekämpfen kann, oder sind Sie dagegen, weil dadurch der Schutz der Privatsphäre zu stark eingeschränkt wird?“

(November 1994, IfD-Umfrage 6008). Ergebnis für Gesamtdeutschland: „Dafür“ 53 %, „Dagegen“ 27 %, „Unentschieden“ 20 %.

Gerade weil Abstimmungsfragen je nach Formulierung zu geradezu entgegengesetzten Ergebnissen führen, müssen, wenn an einer wirklichen Klärung der Einstellung der Bevölkerung gelegen ist, in der Regel mehrere Fragen gestellt werden, um ein Gesamtbild zu gewinnen, und es ist auch die Übersetzung der Programmfragen in Testfragen erforderlich.

Hier spielen die Indikator-Fragen eine wichtige Rolle. In seinem Buch „Regeln der soziologischen Methode“ hat Emile Durkheim (1895) schon zur Jahrhundertwende beschrieben, warum man Indikator-Fragen braucht. Wenn ein Sachverhalt nicht zuverlässig direkt festgestellt werden kann, muß man übergehen auf die Frage nach Sachverhalten, die nicht für sich interessieren, aber von denen man auf den gesuchten Sachverhalt zurückschließen kann. Vertraut mit der Technik von Indikator-Fragen sind Ärzte: „Schmeckt das Pfeifchen noch?“, fragt der Arzt seinen Patienten. Aus der Antwort kann er auf das Stadium einer Krankheit schließen (vgl. Noelle 1963/1976, S. 283).

Oft findet man in einem Fragebogen ganz unbegreifliche Fragen, wie etwa: „Wann haben Sie zuletzt lauthals gelacht?“ Das sind Indikator-Fragen, die zum Beispiel das allgemeine Lebensgefühl, das Befinden des Befragten erforschen, ohne daß der Zweck dafür erkennbar wird.

Zur Übersetzung in Indikator-Fragen gehört auch die Beherrschung von Tests aus der psychologischen Individualdiagnose, Satzergänzungstest, Baumzeichen-Test, Rorschach-Test, Thematic-Apperception-Test (TAT) (vgl. Ring 1992).

Ebenfalls gehört dazu die Kenntnis graphischer Hilfsmittel im Interview, wie auch die Erfindung von Bildern, mit denen der Fragegegenstand anschaulich gemacht wird.

Die Erfindung der Indikator-Fragen als Übersetzung von Forschungsfragen in Testfragen gelingt im Zusammenwirken einer Fragebogenkonferenz besser, als wenn ein Wissenschaftler allein daran arbeitet.

Zum Funktionieren der Fragebogenkonferenz gehört eine bestimmte Ausstattung: ein gut geordnetes Archiv, aus dem mühelos frühere Berichte angefordert werden können, die als Hintergrund für das neue Projekt benötigt werden. Außerdem müssen, mit welcher Technik auch immer, früher entwickelte Fragen zu einem bestimmten Thema, zu einem bestimmten Modell in wenigen Minuten zu beschaffen sein, dann die dazugehörigen Fragebogen, um das frühere Umfeld zu prüfen, und benötigt werden auch die dazugehörigen Code-Pläne mit eingetragenen Gesamtergebnissen („Marginals“), um zu prüfen, wie sich bei früheren Einschaltungen einer Frage die Ergebnisse auf die Antwortkategorien verteilen.

Zudem sollte die Fragebogenkonferenz einen festen Platz haben, einen kleinen Konferenzraum.

Zum Ablauf der Fragebogenkonferenz: Voraussetzung für den Erfolg sind gute Vorarbeit, das Studium der Unterlagen, die zum Projekt und zur Auftragserteilung gehören. Am Anfang der Beratung steht die Aufstellung der Aufgaben der Untersuchung, der Programm-

fragen, eine Beratung der Untersuchungsanlage, der Schwierigkeiten, die zu überwinden sind, der geeigneten Methode. Die Aufzeichnung von Programmfragen, Untersuchungskonzept und Methode ist oft schon Bestandteil des Angebots oder des Projekts.

Es folgt eine gemeinsame Aufstellung von Fragen, teils direkten Fragen, teils Übersetzungen der Programmfragen, das Heranziehen von Modellen, neuen Entwürfen. Die Arbeit verläuft in einem Wechselspiel. Schweigende Arbeit an Texten, dann Diskussion der verschiedenen Entwürfe. In der Regel gibt es dabei große Unterschiede des Ansatzes und der Anschaulichkeit und Lebendigkeit der Sprache.

Man erlebt die Qualitätssteigerung des Fragebogens unmittelbar. Das ist auch der Grund, warum viele Auftraggeber es lieben, ganztägig in der Fragebogenkonferenz mitzuarbeiten. Hierarchie gibt es in der Fragebogenkonferenz nicht, fast immer steht am Schluß der Beratung der einzelnen Fragen oder Fragekomplexe allgemeine Übereinstimmung darüber, was die beste Lösung ist.

Sorge um Zeitverschwendung darf es nicht geben. Nicht selten fordert die Übersetzung einer Programmfrage in die Testfrage eine oder zwei Stunden. Es darf auch keine Bedenken geben, daß viele Entwürfe – etwa 80 % – im Papierkorb landen. Das ist der Preis des optimalen Fragebogens.

Es wird jetzt schon verständlich, wie es kommt, daß 25 % des Budgets einer Untersuchung auf die Fragebogenentwicklung entfallen.

Weitere kostentreibende Faktoren kommen dazu. Jeder Fragebogen wird getestet, und zwar sowohl vom Projektleiter wie von allen Mitgliedern der Fragebogenkonferenz und auch von Mitarbeitern aus vielen Abteilungen des Instituts, aus der Statistischen Abteilung, dem Interviewer-Ressort, der Abteilung für Verschlüsselung offener Fragen, der Datenverarbeitung, der Presseabteilung. Schließlich wird der Fragebogen auch von erfahrenen Interviewern getestet, vor allem bei Fragebogen mit neuer, noch nie bearbeiteter Thematik oder neuer Methode.

Warum müssen die leitenden Wissenschaftler die Fragebogen selbst testen? Es handelt sich bei der Umfrageforschung um einen geistigen Beruf, bei dem genau wie bei Ärzten, Anwälten, Architekten die leitenden Personen den entscheidenden Anteil am Gelingen haben.

Zunächst ist wichtig, daß sich der Umfrageforscher gleichsam freischwimmt. Er muß sich frei machen von den Regeln des konventionellen Umgangs von Menschen untereinander, er muß fremde Menschen ansprechen können, es muß für ihn normal werden, nach persönlichen Dingen zu fragen, über die man im sonstigen Leben nach Konvention keine Fremden fragen würde. Wenn es dem Interviewer, auch dem Wissenschaftler normal geworden ist, solche Fragen zu stellen, erscheint es auch dem Befragten normal. Man kann sagen, daß in der Regel etwa hundert Interviews selbst durchgeführt werden müssen, dann ist man unbefangen geworden bei der Bitte um ein Interview an fremder Wohnungstür.

Das Testen des Fragebogens ist selbst eine Kunst, darum darf der Wissenschaftler das Testen nicht ganz an Interviewer delegieren. Dabei kann eine Systematik wie die von Oksenberg, Cannell und Kalton (1991), woran man schlechte Fragen erkennt, nur nützlich sein. Sie

sagen beispielsweise, daß Rückfragen der Befragten mit der Bitte um Erläuterung immer ein Zeichen sind, daß die Frage überprüft werden muß, ebenso allgemein jede Unterbrechung des Interviews durch den Befragten.

Erwin Scheuch berichtete in seinem Vortrag von einer Studie, bei der systematisch geprüft wurde, ob die Fragen richtig oder falsch verstanden wurden. Die Hälfte aller Fragen sei nicht richtig verstanden worden, sagte er und fügte hinzu: Leider habe es zwischen dem richtigen oder falschen Verständnis der Fragen und anderen Merkmalen wie Bildung, Alter, Geschlecht keine klaren Korrelationen gegeben.

Beim Test bemerkt man, ob die Fragen falsch verstanden werden. Man beobachtet unmittelbar, wie spontan oder langsam und angespannt die Antworten kommen, wie interessiert oder gelangweilt oder auch peinlich berührt der Befragte ist. Man empfindet unmittelbar, daß eine Frage zu lang ist, zu umständlich, zu kompliziert. Gleich im Interview, im Angesicht des Befragten versucht man dann, die einfachere, anschaulichere, kraftvollere Formulierung zu finden. Man löst sich von der Papiersprache und notiert sofort beim Interview, wie die bessere Formulierung lauten könnte. Selbst stundenlange Diskussionen in der Fragebogenkonferenz haben nicht dieses Potential des plötzlich gefundenen Ausdrucks der Umgangssprache.

Ebenso werden Kontext-Effekte oft nur entdeckt, wenn man interviewt. Der Befragte antwortet aus einem Bezugsrahmen heraus, an den niemand in der Fragebogenkonferenz gedacht hatte. George Gallup erzählte mir einmal, nach seiner verheerenden Fehlprognose bei der Präsidentschaftswahl von 1948 habe er sich sofort gedacht: „Wenn ich selbst Probeinterviews gemacht hätte, wäre mir das nicht passiert.“ Von da an, von 1948 an, habe er, solange er das Institut leitete, bei jeder Umfrage selbst Probeinterviews gemacht.

Bestimmte Erlebnisse im Interview vergißt man nie, etwa die Antwort einer Bäuerin auf die Frage „Glauben Sie, daß Goethe ein moralisches Leben geführt hat?“. „Wenn Sie schon so fragen, wird es wohl unmoralisch gewesen sein.“ Oder die Rückfrage nach dem Vorlesen einer Frage: „Kommt danach: Warum?“ Das war der Schlüssel zum Aufdecken der Regel, daß bei Nachfragen nach den Gründen einer Antwort immer weniger konkrete Antworten gegeben werden als ohne Nachfrage, mehr ausweichend: „Da bin ich unentschieden“. Darum werden Trendbeobachtungen gestört, wenn man eine Frage manchmal mit, ein anderes Mal ohne Nachfrage stellt.

Es hat sich bewährt, die Probeinterviews zu institutionalisieren. In Gruppen fahren die Wissenschaftler und die nichtwissenschaftlichen Mitarbeiter in irgendeinen Teil des Ortes, der Stadt, verlassen gemeinsam das Fahrzeug, verteilen sich in verschiedene Richtungen, um Probeinterview-Partner zu suchen, ihre Interviews zu machen, und treffen sich zur vereinbarten Zeit zwei, drei Stunden später wieder.

Die Fragebogen-Korrekturarbeiten sollten sich so schnell wie möglich an die Probeinterviews anschließen. Zuerst trägt jeder in den Korrekturbogen seine Erfahrungen, Beobachtungen, Korrekturvorschläge ein. Die Konferenz mit der Diskussion schließt sich an. Dabei werden nach Bedarf die Ergebnisse einzelner Fragen aus den Probeinterviews mit Strichliste ausgezählt, zum Teil zeigt schon die Verteilung der Antworten, daß eine Frage einseitig ist oder falsch verstanden wurde. Der Kostenaufwand ist gerechtfertigt.

Natürlich ist es auch sinnvoll, die in der Feldarbeit besonders erfahrenen Interviewer nach ihren Erfahrungen zu fragen, um ihre Kritik zu bitten und in jedem Fall das Schreiben eines Erfahrungsberichts zur Regel zu machen, der dann nicht nur im Interviewer-Ressort gelesen (und gegebenenfalls beantwortet!) wird, sondern auch von den Mitgliedern der Fragebogenkonferenz. Peter Mohler fragte in der Diskussion: „Gibt es gute Pretest-Interviewer?“ Erwin Scheuch antwortete: „Nehmen Sie die besten!“ Wenn die Wissenschaftler selbst testen, sollten sie keine Spezialtruppe von guten Pretest-Interviewern aufbauen. Mit der Information an die Interviewer „Das ist ein Pretest!“ geht etwas von der Normalität verloren, auf die es bei den Reaktionen der Interviewer gerade ankommt. Außerdem muß man immer an die Warnung des ersten Direktors des britischen Gallup-Instituts, Dr. Henry Durant, denken: „Holen Sie die Interviewer nie zusammen, um sie zu schulen oder ihre Erfahrungen mit ihnen zu diskutieren. Interviewer zusammen: Das ist wie eine Diebesküche („thief’s kitchen“ heißt es auch im Englischen!), sie tauschen nur Tricks aus!“

Schon Mohler wies darauf hin, daß je nach dem Ausmaß der erforderlichen Korrekturen nicht nur ein Pretest, sondern zwei oder drei nacheinander geschaltete Pretests erforderlich sein können. „Never written a perfect questionnaire on the first trial“ wurde in der Diskussion gesagt.

Hier verlasse ich mein Thema „Fragebogenkonferenz“ und füge noch einige Überlegungen zum Pretest, dem zweiten Hauptthema dieser Arbeitstagung, bei.

Der Begriff Pretest wird sinnvollerweise weiter gefaßt als gemeint ist, wenn man nur von der Prüfung eines voll entwickelten strukturierten Fragebogens spricht. Die sogenannten „qualitativen“ Untersuchungen mit unstrukturierten oder halbstrukturierten Fragebogen („Tiefeninterview“), Fall-Studien, Explorationen, „fokussierten Interviews“ oder Gruppendiskussionen bilden im Ablauf einer Untersuchung nicht Endpunkte, obwohl sie leichtfertig oft so genutzt werden. Ihrer Funktion nach sind sie Vorstudien zur Denkweise, Reaktionsweise der Zielgruppe zu bestimmten Forschungsthemen, Assoziationen, zum Sprachgebrauch. Im strukturierten Fragebogen findet sich sinnvollerweise die gegabelte Befragung als Pretest. Michaela Wänke, Universität Heidelberg, Psychologisches Institut, formulierte: „Die gegabelte Befragung möchte ich als den anderen Pretest bezeichnen.“

Höhnisch ist manchmal gesagt worden, die meisten Abhandlungen über die Fragebogen-Methode läßen sich, als trügen sie den Titel: „Die Kunst, eine Frage zu stellen“.

Tatsächlich muß der Pretest nicht nur einzelne Fragen betreffen, sondern Fragemodelle, die mehrere Fragen umfassen, müssen getestet werden. Ganze Untersuchungskonzepte, wie etwa die Internationale Werte-Studie – zuerst 1981 (Allensbacher Archiv 1981) und dann 1990/91 (Allensbacher Archiv 1990/91; vgl. Nolle-Neumann, Köcher 1987) – die schließlich erfolgreich in fast dreißig Ländern durchgeführt wurde, mußten als Ganzes getestet werden, die Allensbacher Markt- und Werbeträger-Analyse AWA seit 1959, seit 1992 rund 20 000 Interviews pro Jahr, die ganze Dramaturgie des Fragebogens. Dabei klärt sich auch, ob man sich dem Standpunkt von Erwin Scheuch anschließen soll: „Wir haben gute Erfahrungen mit demographischen Fragen am Anfang gemacht, zusammen mit der Angabe ‚Eine Untersuchung der Universität Köln‘ wirkt das amtlich oder halbamtlich“. Oder der entgegengesetzten Auffassung von Akkerboom: „*Amtlichkeit* ist nicht gerade etwas, womit wir die Befragten zu erfreuen versuchen.“ Das Ziel der ersten Fragen im Interview ist, das Interesse

des Befragten zu wecken, ihn zu motivieren, das Interview zu geben und durchzuhalten. Ob das gelingt, läßt sich im Pretest beobachten. Viele Fragen weisen sich schon dadurch als ungeeignet aus, daß die meisten Befragten sie mit „Nein“ beantworten. Ein „Nein“ am Anfang des Interviews ist grundsätzlich kein guter Beginn, Fragen, die die meisten verneinen, sind keine guten Eisbrecher-Fragen.

Die Qualität einer Untersuchung auf neuem Feld wird deutlich gesteigert, wenn die Untersuchung als Mehr-Stufen-Modell angelegt wird. Nur die verbreitete Gleichgültigkeit gegen Qualitätsunterschiede in der empirischen Sozialforschung erklärt, daß das Mehr-Stufen-Modell nicht längst bei schwierigen Aufgaben selbstverständlich ist. Verhältnismäßig häufig begegnet man noch der *zweistufigen* Untersuchung, am Anfang Gruppendiskussion, fokussierte Interviews, unstrukturierte „Tiefeninterviews“ und, auf dieser Vorstufe aufbauend, dann der strukturierte Fragebogen der Hauptstufe.

Kaum entdeckt ist der große Vorzug einer weiteren Zwischenstufe, bei der mit ganz oder halb strukturierten Fragebogen ein Pretest mit etwa 200 Interviews durchgeführt, ausgezählt und analysiert wird. Der Grund für die Empfehlung der Zwischenstufe ist, daß keine noch so intensive Fall-Studie, kein noch so langgedehntes Tiefeninterview Sachverhalte zum Vorschein bringen kann, die dem Befragten nicht bewußt sind und über die er darum auch nicht sprechen kann. Gewiß gibt es Indikatoren in dem schon beschriebenen Sinn, Freudsche Versprecher gehören wahrscheinlich dazu. Aber die Interpretationen sind weitgehend subjektiv, der Nachweischarakter des „Wiederholbar, Überprüfbar“ fehlt.

Beim Übergang vom Individuum der Fall-Studie auf das Aggregat eines Miniatur-Querschnitts lassen sich Korrelationen finden, die dem einzelnen unbewußt sind. Auf der Makro-Ebene, auch wenn sie noch klein ist, kann man nach Zusammenhängen suchen, die die Anlage des endgültigen Fragebogens beeinflussen werden. 200 Interviews erlauben auch schon, ein kontrolliertes Feldexperiment durchzuführen, mit je 100 Interviews in gegabelter Befragung, besonders geeignet, um Unbewußtes hervortreten zu lassen. Es kommt hinzu, was zuvor schon über die Strichliste beim Auszählen einer kleinen Zahl von Probe-Interviews gesagt wurde. Hier gilt erst recht: Die erste Übersicht, wie die Antworten auf die geplanten Fragen streuen werden, zeigt oft bereits, daß eine Frage nicht so funktioniert, wie erwartet und gewünscht. Ein besonderes Alarmzeichen ist es immer, wenn viele Befragte unentschieden bleiben.

Daß die Zwischenstufe eines strukturierten Pretests mit 200 Interviews so wenig verbreitet ist, hängt auch mit einem Dogma zusammen, das die empirische Sozialforschung heute noch wie einen gefesselten Riesen erscheinen läßt. Dem Dogma nach wird eine Umfrage abgeleitet aus einer Forschungshypothese, die verifiziert oder falsifiziert werden soll.

Ganz vernachlässigt ist das Entdeckungspotential empirischer Forschung, das Hervortreten unerwarteter Reaktionen, die Überraschung. Das Entdeckungspotential der Umfrageforschung läßt sich vorsätzlich durch Gebrauch geeigneter Methoden steigern (Noelle-Neumann 1989). Entdeckung bedeutet Korrektur falscher Annahmen und Anregung neuer Forschung – ganz wie in der Naturwissenschaft.

Hier steht man vor dem großen, fast noch nicht bearbeiteten Feld: Qualitätskriterien in der Umfrageforschung. Kein anderer Faktor beeinflusst die Ergebnisse der Umfrage so sehr wie der Fragebogen. Also muß die Formulierung von Qualitätskriterien beim Fragebogen beginnen, sich auf den Fragebogen konzentrieren.

Literaturhinweise

Allensbacher Archiv (1981): IfD-Umfrage 1295.

Allensbacher Archiv (1990/91): IfD-Umfragen 2287, 3201.

Bradburn, N., Sudman, S. u. a. (1979): Improving Interviewing Method and Questionnaire Design, San Francisco.

Cantril, H. (1947): Gauging Public Opinion, Princeton.

Durkheim, E. (1895): Les règles de la méthode sociologique. (Deutsch: Regeln der soziologischen Methode, Neuwied, 1961.)

Noelle, E. (1963/1976): Umfragen in der Massengesellschaft. Einführung in die Methoden der Demoskopie, Reinbek bei Hamburg 1963, 7. Auflage 1976, Band 177.

Noelle-Neumann, E. (1970): Wanted: Rules for Wording Structured Questionnaires, in: Public Opinion Quarterly, 34, S. 191 – 201.

Noelle-Neumann, E. (1974): Draft on an International Research Program with Regard to Questionnaire Methodology, in: Allensbacher Archiv, Vortrag am 31. Mai 1974, Jahreskongreß der American Association of Public Opinion Research (AAPOR) in Lake George/USA.

Noelle-Neumann, E. (1989): Kreativität und Nüchternheit sind keine Gegensätze. Die Umfragemethode ist gut für Registrieren und Testen, aber warum wird das Entdeckungspotential nicht systematisch entwickelt?, in ESOMAR (Hrsg.): The 42nd ESOMAR Marketing Research Congress, Stockholm, 3rd – 7th September 1989, Amsterdam, S. 377 – 400 (deutsch), S. 401 – 426 (englisch).

Noelle-Neumann E., Köcher, R. (1987): Die verletzte Nation. Über den Versuch der Deutschen, ihren Charakter zu ändern, Stuttgart.

Oksenberg, L., Cannel, C., Kalton, G. (1991): New Strategies in Pretesting Survey Questions, in: Journal of Official Statistics, Nr. 3, 7. Jg., S. 349 – 365.

Payne, S. L. (1951): The Art of Asking Questions, Princeton.

Ring, E. (1992): Signale der Gesellschaft. Psychologische Diagnostik in der Umfrageforschung, Göttingen, Stuttgart.

Schumann, H., Presser, S. (1981): Questions and Answers in Attitude Surveys. Experiments on Question Form, Wording, and Context, New York, London, Toronto, Sydney, San Francisco.

Sudman, S., Bradburn, N. (1974): Response Effects in Surveys. A Review and Synthesis, Chicago.

Sudman, S., Bradburn, N. (1982): Asking Questions. A Practical Guide to Questionnaire Design, San Francisco.

Labor für die Entwicklung und den Test von Erhebungsinstrumenten

1 Das CBS-Labor für die empirische Erforschung von Fragebogen

1.1 Gründung des CBS-Labors

Im Jahr 1992 hat das „Centraal Bureau voor de Statistiek (CBS)“, das niederländische Zentralamt für Statistik, begonnen, ein sogenanntes „Vragenlab“ (Fragebogenlabor) aufzubauen, ein Labor für die Entwicklung und den Test von Erhebungsinstrumenten. Hiermit wurde an vielversprechende Entwicklungen in der Erhebungsmethodologie angeknüpft, die in den achtziger Jahren mit der CASM-Bewegung (Cognitive Aspects of Survey Methodology) in den Vereinigten Staaten ihren Ursprung hatten.¹⁾ Die vermeintliche Einfachheit des Frage-und-Antwort-Prozesses hat sowohl die Datenanalytiker als auch die Erhebungsspezialisten geblendet, sie die spezifischen Fehlerquellen nicht sehen lassen. Durch CASM versuchten Kognitionspsychologie und Erhebungsmethodologie, gemeinsame Wege zum besseren Verständnis der Informationsverarbeitungsabläufe im Frage-und-Antwort-Prozeß zu finden. Das bedeutete eine *Befragten- (und Interviewer-) bezogene Erneuerung in der Erforschung der Qualität von Fragebogen*. Die CASM-Bewegung machte mit neuen Methoden darauf aufmerksam, daß bei der Beantwortung einer Frage noch mehr „abläuft“ oder „schiefläuft“ als ein schlichter Stimulus-Response-Prozeß. Was im Frage-und-Antwort-Prozeß abläuft, kann und soll an erster Stelle empirisch erfaßt werden. Der eigentliche Entschluß zur Gründung eines CBS-Labors beruhte eher auf einer Analyse der erfolgreichen pragmatischen Arbeitsweise eines Fragebogenlabors in Kanada als auf wissenschaftlichen Erkenntnissen durch die CASM-Bewegung.²⁾

Probleme im Fragebogen sollten möglichst vor der eigentlichen Erhebung erkannt und durch Änderungen unterbunden werden. So rückt der Anteil des Fragebogenlabors an der Qualitätssicherung des ganzen Erhebungsprozesses in den Vordergrund. Die Begründung der benutzten Testverfahren ist praktischer Natur und beruht höchstens indirekt auf der ohnehin vielfältigen Grundlagenforschung (in die neben kognitionspsychologischen Ansätzen auch Beiträge aus Linguistik, Kommunikationstheorie und Sozialpsychologie einbezogen sind).

1.2 Pretests und Posttests durch Meta-Befragung

Unter einem empirischen Qualitätstest eines Fragebogens wird im CBS-Labor ein Testprogramm verstanden, bei dem, aus verschiedenen Blickwinkeln betrachtet (Befragter, Interviewer, unabhängiger Beobachter), überprüft wird, inwiefern mit einem Fragebogen die für eine Fragestellung erforderlichen Daten erhoben werden können. Wenn dieser Test vor der

*) Dr. Hans Akkerboom, Centraal Bureau voor de Statistiek, Heerlen, Niederlande.

1) Vgl. Tanur, J. M., Fienberg, S. E.: Cognitive Aspects of Surveys: Yesterday, Today and Tomorrow, Journal of Official Statistics, Band 8 (1), 1992.

2) Vgl. Gower, A. R.: The Questionnaire Design Resource Centre's Role in Questionnaire Research and Development at Statistics Canada, 48th Session of the International Statistical Institute, Band 3, 1991, S. 58 – 59.

eigentlichen Erhebung oder vor einer Versuchserhebung oder Generalprobe stattfindet, dann kann manche Unstimmigkeit im Fragebogen rechtzeitig entdeckt und behoben werden. Das CBS schlägt folgende Definition für den Begriff „Pretest“ vor:

Pretest = empirische Forschung zur gegenseitigen Abstimmung von Informationsangebot und -nachfrage bei der Entwicklung eines Fragebogens.

In dieser Definition deutet „gegenseitige Abstimmung“ darauf hin, daß inhaltliche Konzepte und praktische Verfahren bei der Erhebungsorganisation mit Denk- und Handlungsmustern der Befragten in Einklang gebracht werden müssen. Sowohl inhaltliche Experten wie Befragte sollten daher in das Pretestprogramm einbezogen werden.

Beim Pretest geht man von einem ersten Konzept des Fragebogens (und der Datenerfassungsverfahren) aus: Eine *Entwurfsphase* geht dem Pretest voraus. Bei kontinuierlichen oder wiederholten Erhebungen ist dieses „vor“ und „nach“ nicht immer klar zu unterscheiden. Bei anderen Erhebungen ist die Möglichkeit eines vorausgehenden Tests gar nicht gegeben. In solchen Fällen kann, neben einer quantitativen Bewertung der Datenqualität, ein sogenannter „Posttest“ dennoch sinnvoll sein. Das Ziel eines Posttests ist es nicht, eventuelle Meßfehler (Nichtstichprobenfehler) zu erfassen, sondern es geht darum, beim Befragten dessen Gründe für sein Antwortverhalten oder seine Antwortverweigerung zu erforschen. Ein Posttest kann also sinnvoll sein zum besseren Verständnis der quantitativen Analyse und im Hinblick auf zukünftige Anpassungen des Fragebogens.

Die Versuchsmethoden und -techniken eines Labortests, im Unterschied zu einer rein quantitativen Bewertung (wie z. B. in einer Pilotstudie), gehen alle von der Annahme aus, daß die Antworten auf die Fragen nicht aussagekräftig genug sind, um die Qualität des Fragebogens zu beurteilen. Man will darüber hinaus „qualitativ“ erfahren, wie der Befragte zu seinen Antworten kommt, was er dabei denkt, welche Daten ihm zur Verfügung stehen, was er von den Fragen hält usw. Die gebräuchlichsten Laborverfahren mit Versuchspersonen benutzen „Meta-Befragungsmethoden“:

1. *Kognitive Interviews* (Einzelexplorationen) mit Meta-Befragungstechniken wie Think-Aloud-Protokolle, Paraphrasing, Erläuterungsfragen.
2. *Gruppendiskussionen* (Mehrfachexplorationen), z. B. im Anschluß an die schriftliche oder mündliche Befragung einzelner Gruppenmitglieder.
3. *Probing* (Rückfrageverfahren vor allem aufgrund von vorgegebenen Checklisten, gezieltes Nachfragen).

Das Aufschlußreiche, aber auch Schwierige der Meta-Befragungstechniken besteht darin, daß, außer einer „schlichten“ Antwort auf eine „schlichte“ Frage, vom Befragten auch eine Mitwirkung in Form von Selbstbeobachtung verlangt wird. Neben direkter Meta-Befragung gibt es indirekte qualitative Testmethoden, wobei die Aufgaben der Versuchspersonen sich auf die eines normalen Befragten beschränken. Hier braucht man einen Beobachter (Interviewer oder Forscher), der Aspekte des Frage- und Antwortverhaltens während des Interviews aufzeichnet: Checklisten oder Bewertungsfragen für den Interviewer, Gruppendiskussionen mit Interviewern, Interaktionskodierung usw. In der Entwurfsphase sind auch

Gruppendiskussionen hilfreich, neben der Fragebogenkonferenz, der Beratung durch Experten und der einschlägigen methodologischen Literatur.

1.3 Entwicklung des CBS-Labors

1992 und 1993 ging es beim Aufbau des Labors darum zu lernen, wie

- *Versuchspersonen* rekrutiert, registriert und nach Bedarf für Tests eingeladen werden können (Rekrutierungssystem und Datenbank mit freiwilligen Testteilnehmern; finanzielle oder sonstige Belohnung),
- *Labormitarbeiter* sowohl Meta-Befragungstechniken als auch standardisierte Befragungstechniken anwenden sollen (Training und Supervision),
- *Pretestprotokolle* formuliert und standardisiert werden, um schnell und effektiv Ergebnisse bereitzustellen.

Wichtiger Ausgangspunkt in dieser Anfangsphase, in der ausschließlich Personenfragebogen getestet wurden, war, daß die Befragungsbedingungen der eigentlichen Erhebung, wo immer möglich, in die Laborsituation übernommen wurden. Auch wenn nur einzelne Fragen erprobt wurden, dann sollte das im gegebenen Fragebogenkontext (Aufbau und Reihenfolge der Fragen) geschehen. Da vom CBS die meisten Personenbefragungen computerunterstützt durchgeführt werden, wurden kognitive Interviews auf Basis solcher Verfahren aufgebaut. Für einen Pretest werden meistens insgesamt etwa 20 bis 40 Versuchspersonen ausgewählt. Hierbei wird eine gewisse Ausgewogenheit im Hinblick auf Ausbildungsniveau, Alter, Geschlecht oder auch erhebungsspezifischen Merkmalen angestrebt.

Zunächst hat das CBS-Labor sich auf kognitive Interviews konzentriert, da diese am einfachsten an standardisierte Interviews anknüpfen. Der nächste Schritt war, daß einige Labormitarbeiter sich die Fähigkeiten eines Diskussionsleiters aneigneten. Seit 1994 werden Gruppendiskussionen im Wechsel mit kognitiven Interviews in gemischte Pretestprogramme aufgenommen. Seitdem wurden und werden andere Erweiterungen der Laborarbeit entwickelt: (1) Tests von Fragebogen für Unternehmen und Behörden³⁾, (2) Gruppendiskussionen mit Interviewer, (3) Gruppendiskussionen zum Fragebogenentwurf, (4) Beratung von Fragebogenkonzepten, (5) Kurse zur Fragebogengestaltung und (6) Anwendung von computerunterstützter Meta-Befragung bei Qualitätstests unter realistischen Befragungsbedingungen („Feldlabor“). Momentan werden etwa fünf bis zehn Fragebogen pro Jahr einem Pretest unterzogen.

2 Beiträge des CBS-Labors zur Qualitätssicherung von Erhebungen

2.1 Meta-Befragung und Pretestmanagement

Das CBS-Labor begann seine Arbeit 1992 mit zwei Mitarbeitern aus der Abteilung für Statistische Methoden. Seit Mitte 1994 gehört das Labor zu einer neuen Abteilung für Beobachtungsmethodologie und verfügt mittlerweile über etwa sechs Mitarbeiter. Diese sind in der Mehrzahl sozialwissenschaftlich ausgebildet, mit einer Spezialisierung in empirischen Forschungsmethoden und -techniken. Sie haben unterschiedliche Erfahrungen mit Daten-

3) Vgl. Gower, A. R.: Questionnaire Design for Business Surveys, Survey Methodology (20), 1994, S. 125 - 136.

erfassungsmethoden, sei es aus der Praxis oder aus dem Entwurf von „elektronischen“ Fragebogen.

Aus den bisherigen Erfahrungen lassen sich zwei gewünschte *Profile* von Eigenschaften und Fähigkeiten von Fragebogenlabormitarbeitern ableiten:

Übersicht 1 Anforderungsprofile von Fragebogenlabormitarbeitern

<i>Meta-Befragungsexperten (Pretestausführung)</i>	<i>Pretestprogrammexperten (Pretestmanagement)</i>
1. Kommunikative Flexibilität	1. Kommunikative Flexibilität
2. (Meta-)Befragungstechniken	2. Pretestprotokolle und -methoden
3. Sachkenntnisse bzgl. Rahmenbedingungen, Erhebungsthemen und Ziel	3. Sachkenntnisse bzgl. Rahmenbedingungen, Erhebungsthemen und Ziel
4. Entwurf/Analyse des Fragebogenkonzepts	4. Entwurf/Analyse des Fragebogenkonzepts
5. Auswertung mündlicher Daten	5. Interpretation der Meßfehler

Kommunikative Flexibilität (und Effizienz) stellt vorrangig eine Fähigkeit von Fragebogenlabormitarbeitern dar, ist aber ohne Erfüllung der übrigen Qualifikationsmerkmale bedeutungslos. Die Meta-Befragungsexperten sollten flexibel auf unvorhersehbare Verhaltensweisen von Versuchspersonen reagieren können (es gibt z. B. die „Angeber“, „Außenseiter“, „Stürmer“ und „Jasager“). Vergleichbares gilt für den Pretestmanager, der oft eine Vermittlerrolle spielt zwischen inhaltlichen Experten und den Befragten, die letztlich die Information geben sollen. Die Meta-Befragungsexperten haben oft die Aufgabe, den Wünschen der Auftraggeber oder inhaltlicher Experten entgegenzutreten. Einerseits lassen sich bestimmte Fragestellungen nicht in der vorgesehenen Form erheben, andererseits kann eine professionelle Meta-Befragung daran scheitern, daß Experten durch direkte Beteiligung „mitmischen“ wollen. Allerdings können bei der Beobachtung von einigen Pretestinterviews durch den Auftraggeber, z. B. mittels Einweg-Scheibe oder Videoaufzeichnung, Beantwortungsprobleme sehr deutlich vermittelt werden.

Die Meta-Befragungstechniken erfordern besondere Fähigkeiten zur Motivation und Lenkung der Versuchsperson(en). Es scheint, daß kognitive und motivationale Schwierigkeiten bei der Beantwortung von Fragen weitaus häufiger sind und sehr viel ernstere Konsequenzen haben, als man üblicherweise annimmt. Dies kann die Durchführung von Meta-Interviews erleichtern, weil auf den Befragten besser eingegangen wird, als es in einem standardisierten Interview möglich wäre. Die geforderte Selbstbeobachtung kann dagegen dem Befragten Schwierigkeiten bereiten, die es im üblichen Interview nicht gibt.

Für aufeinanderfolgende Pretests sollte der Projektmanager fähig sein, Änderungen des elektronischen Fragebogens bzw. Testprotokolls selbst zu programmieren (in der BLAISE-Software). Natürlich überschneiden sich die erforderlichen Qualifikationen. Der Ausgewogenheit der einzelnen Fähigkeiten kommt dann besondere Bedeutung zu, wenn man versucht,

Labormethoden bei einer größeren Zahl von Befragten im eigenen Haus oder am eigenen Arbeitsplatz durchzuführen. Für die Feldlaborarbeit benötigt man qualifizierte Interviewer, die Meta-Befragungstechniken beherrschen sollten.

2.2 Neugestaltung des CBS und seiner Datenerfassungsverfahren

Pretests oder Posttests braucht man, wenn eine Erhebung qualitäts- und kostenbewußt durchgeführt werden soll. Der mögliche Gewinn des Pretests beruht allgemein gesagt darauf, daß man (1) vorbereitende Diskussionen zum Entwurf eines Fragebogens auf der Basis empirischer Daten beschleunigen kann, und (2) Qualitätsuntersuchungen während und nach der eigentlichen Erhebung durch kleinere, frühzeitige Pretests ersetzt werden können. Die meisten Meßfehlerrisiken können so mit viel geringerem Aufwand (Kosten, Zeit) aufgedeckt werden.

Ein wichtiger Impuls für den Pretest formte die 1993 eingesetzte Neugestaltung des CBS unter dem Motto „TEMPO“:

Übersicht 2 Auswirkungen der Neugestaltung von CBS auf den Datenerhebungsprozeß

<i>Motto</i>	<i>Konsequenzen für Fragebogen/ Datenerfassung</i>
„Tijdig“ = rechtzeitig	– Vereinfachung und Integration (auch zur Verringerung der Umfragebelastung)
Effizient	– Mehrfache Erhebungsmodi (auch zur Erhöhung der Aus- schöpfungsquote)
Modern	– Computerunterstützte Befragung oder elektronische Datenerfassung
Professionell	– Datenlieferanten- und Kundenorientierung, Erhebungsmethodologie
„Onafhankelijk“ = unabhängig	– Sichtbare Qualität

Die Neugestaltung führte zu einer organisatorischen Einteilung gemäß den verschiedenen Phasen des Erhebungsprozesses statt einer traditionellen Einteilung nach inhaltlichen Bereichen des statistischen Programms. Damit wurde die gesamte Datenerfassung mit all ihren Arbeitsvorgängen in einer Abteilung konzentriert, einschließlich der Erhebungsmethodologie. Diese Umorganisation löste eine Neugestaltung der statistischen Programme und der Erhebungsprozesse aus. Die Kombination von Datenerfassung aus administrativen Registern und Datenerfassung mittels mehrerer Erhebungsmodi führt beispielsweise zu grundlegenden Fragen der Datenqualität. Die Datenlieferanten- und Kundenorientierung als ein Ziel der Neugestaltung von CBS ist ein wichtiges Argument für den Pretest, wichtiger als die Fortschritte in der Grundlagenforschung seit der CASM-Bewegung.

In der Grundlagenforschung werden allgemeine Erkenntnisse über den Frage-Antwort-Prozeß gewonnen, um Methoden zu entwickeln, mit denen die Gültigkeit von Antworten verbessert werden kann und mit denen striktere Anforderungen an den Meßprozeß erfüllt werden können. Dies kann am ehesten durch eine Verbesserung der Techniken zur Fragebogenkonstruktion erreicht werden. Von der Kognitionspsychologie oder anderen Disziplinen wird hier ein Beitrag zur Professionalisierung der Fragebogenerforschung erwartet. Testergebnisse wären dadurch einfacher und systematischer zu analysieren und in Änderungsvorschläge umzusetzen. Gewisse Faustregeln des Fragebogenentwurfs wären besser zu begründen. Trotzdem müssen auch bewährte Fragestellungen in jedem neuen Kontext neu erprobt werden.

Wissenschaftliche Beiträge zur Beobachtungsmethodologie sind vielleicht am meisten gefragt, wenn es darum geht, die folgenden Fragen zu beantworten:

1. Wie belegt man, abgesehen von konsistenten Ergebnissen aus unabhängigen Quellen, die *Validität und Zuverlässigkeit der Meta-Befragung* (Kognitive Voraussetzungen; Verzerrung der Meta-Information durch ungenügend sachgerechte Selbstbeobachtung usw.)? Welcher Pretestumfang genügt für die Aufdeckung der wichtigsten Meßfehlerisiken?
2. Wie laufen Erinnerungs- und Frage-und-Antwort-Prozesse ab? Wie können diese Prozesse optimiert werden?
3. Nach welchen Qualitätsmerkmalen und mit welchen Methoden bestimmt man die *Machbarkeit* einer Frage oder gar eines Erhebungsziels?

3 **Schlußfolgerung**

Das CBS-Labor erfüllt immer weitergehende Aufgaben in der Qualitätssicherung des Erhebungsprozesses. Labortests basieren auf der Feststellung, daß es „angemessene, präzise und vollständige Daten ohne Mitwirkung des Befragten nicht gibt“. In der Phase des Fragebogenentwurfs geht es darum, den statistischen Programmfragen die Frage „Welche Erhebungsthemen sind relevant und meßbar aus der Sicht des Befragten?“ gegenüberzustellen. In der Pretestphase wird der Konzeptfragebogen nach folgenden Gesichtspunkten erprobt: „Wie verhält sich die Interpretation der Fragen und Antworten des Befragten zu der beabsichtigten Fragestellung? Ist der Befragte fähig und gewillt, die benötigten Daten zu geben?“ Bei nachfolgenden Qualitätsuntersuchungen will man noch wissen: „Auf welche Meßfehlerisiken ist weiterhin zu achten?“ Fragebogentests können zuerst effizient im Labor durchgeführt werden. Danach sollten sie auch unter realistischeren Befragungsbedingungen stattfinden; „Labor“ bedeutet dann eher, daß man dem Befragten da, wo er wohnt oder arbeitet Meta-Information abverlangt, statt nur einfach eine Antwort auf eine statistisch motivierte Frage zu erhalten. Zur professionellen Erforschung von Fragebogen braucht man die Weiterentwicklung einer *Disziplin der Beobachtungsmethodologie*. Pretests stellen dabei ein wichtiges, pragmatisches Werkzeug dar.

Pretest zur Erhebung „Neukonzeption der laufenden Wirtschaftsrechnungen“

Vorbemerkung

Die Wirtschaftsrechnungen sind in den vergangenen Jahren zunehmender Kritik ausgesetzt gewesen. Sie basieren auf den monatlichen Anschreibungen von rund 2 000 ausgewählten Haushalten in Haushaltsbüchern. Einbezogen werden Zwei-Personen-Haushalte von Renten- und Sozialhilfeempfängern mit geringem Einkommen, Vier-Personen-Haushalte von Angestellten und Arbeitern mit mittlerem Einkommen und Vier-Personen-Haushalte von Beamten und Angestellten mit höherem Einkommen. Nachgewiesen wird unter anderem das monatliche Haushaltsbudget, die Einnahmen der Haushalte nach Quellen, ihre Ausgaben nach Arten und nach Verwendungszwecken für den Privaten Verbrauch sowie die Ausstattung mit langlebigen Gebrauchsgütern. Der bewußt gewählte geringe Repräsentationsgrad der laufenden Wirtschaftsrechnungen – geringer Stichprobenumfang und Begrenzung auf nur drei Haushaltstypen – schränkt die Verwendbarkeit der Ergebnisse stark ein. Wünschenswert wäre statt dessen, die laufenden Wirtschaftsrechnungen als ein Instrument zu nutzen, das die Datenlücke zwischen den einzelnen Einkommens- und Verbrauchsstichproben (EVS) mit statistisch ausreichend gesicherten Daten füllt.

Die EVS werden fünfjährlich bei über 50 000 Haushalten aller Bevölkerungsschichten durchgeführt. Sie erbringen Angaben über die Zusammensetzung sowie die wirtschaftlichen und sozialen Verhältnisse der Haushalte, ihre Einnahmen nach Quellen und Angaben nach Arten und Verwendungszweck. Ferner werden die Vermögensformen, -bestände und Schulden dargestellt. Das Programm beider Erhebungen soll aufeinander abgestimmt werden, um aus ihnen ein Gesamtsystem der Wirtschaftsrechnungen zu entwickeln.

Im Rahmen der Neukonzeption der laufenden Wirtschaftsrechnungen (LWR) wurden die neu entwickelten Erhebungsunterlagen (Einführungsinterview und Haushaltsbuch) mit Hilfe eines Pretests optimiert. Bei der Erprobung der Erhebungspapiere sollten, um eine hohe Datenqualität sicherzustellen, nicht nur traditionelle Formen des Beobachtungspretests zum Einsatz kommen. In die Vorarbeiten wurde das gesamte Spektrum der anerkannten sozialwissenschaftlichen Pretestverfahren einbezogen, um eine Grundlage zur Durchführung eines methodisch breiter angelegten und sich moderner Techniken bedienenden Pretestverfahrens zu schaffen. Mit der Durchführung des Pretests beauftragte das Statistische Bundesamt das Zentrum für Umfragen, Methoden und Analysen (ZUMA) in Mannheim.

1 Konzeption des Pretests

Mit Hilfe des Pretests sollte in erster Linie die Praktikabilität der Erhebungsunterlagen vornehmlich unter folgenden Gesichtspunkten untersucht werden: Länge des Einführungsinterviews und zeitlicher Aufwand für das Ausfüllen des Haushaltsbuches, Verständnis der Fra-

^{*)} Dr. Manfred Ehling, Statistisches Bundesamt, Wiesbaden.
Rolf Porst, Zentrum für Umfragen, Methoden und Analysen (ZUMA), Mannheim.

gen, Erläuterungen und Hinweise, Reihenfolge der Erhebungsinhalte beider Instrumente, das Layout der Erhebungsunterlagen sowie die Motivation zur Teilnahme. Daneben sollten auch Erkenntnisse im Hinblick auf die Akzeptanz der Neukonzeption der laufenden Wirtschaftsrechnungen und den Einsatz von DV-gestützten Erhebungsinstrumenten gewonnen werden.

Für diesen geplanten Instrumentenpretest wurde daher folgendes Design vorgesehen: In der *ersten Phase* werden acht Haushalte per Quote ausgewählt. Die Bezugspersonen (Haupteinkommensbezieher) in diesen Haushalten füllen das Einführungsinterview in Gegenwart eines Interviewers (wissenschaftliche Mitarbeiter der Feldabteilung von ZUMA) aus. Nach dem Ausfüllen werden fokussierte Intensivinterviews zum Einführungsinterview durchgeführt und durch Tonbandaufzeichnung festgehalten.

Die Bezugspersonen sollen dann zunächst 7 Tage lang das Haushaltsbuch ausfüllen. Am siebten Tag geschieht dies in Gegenwart eines Interviewers, der dabei schon Kommentare der Bezugsperson registrieren kann. Am Ende der Ausfüllphase kommt dann erneut ein fokussiertes Intensivinterview zum Einsatz, diesmal bezogen auf das Haushaltsbuch, das ebenfalls per Tonband aufgezeichnet wird. Alle Bänder werden gemeinsam von Mitarbeitern von ZUMA und des Statistischen Bundesamtes abgehört und interpretativ ausgewertet.

Im Rahmen von zwei Fragebogenkonferenzen sollen anschließend die Entwürfe der Erhebungspapiere diskutiert werden. Bei ZUMA besprechen in dieser *zweiten Phase* wissenschaftliche Mitarbeiter der Feldabteilung und erfahrene Interviewer, die die Bogen selbst bearbeitet haben, die Stärken und Schwächen der Instrumente. In der zweiten Fragebogenkonferenz – im Statistischen Bundesamt – werten unabhängig davon die zuständigen Bearbeiter unter Berücksichtigung der Erfahrungen aus den Tonbandprotokollen die Beobachtungen aus der ersten Pretestphase aus. Die Resultate der ersten und zweiten Phase fließen in eine Überarbeitung der Erhebungspapiere ein.

In der *dritten Phase* werden mit den überarbeiteten Instrumenten danach von erfahrenen ZUMA-Interviewern in 36 Haushalten Beobachtungspretests durchgeführt, die durch Probing-Verfahren (comprehension probes, information retrieval probes und confidence ratings) ergänzt werden. Jeder Haushalt erhält zunächst den Fragebogen des Einführungsinterviews, die Bezugsperson füllt ihn aus und wird dann mit den genannten Verfahren zum Einführungsinterview befragt. Eine zweite Befragung erfolgt eine Woche nach dieser Erstbefragung und konzentriert sich, ebenso wie die dritte Befragung (2 Wochen nach der Erstbefragung) beim Abholen der Instrumente, auf das Haushaltsbuch. Die Befragten haben dann insgesamt 14 Tage lang ein Haushaltsbuch zu führen.

Kernpunkt der Konzeption des Instrumentenpretests bildet ein *Pretestverfahren mit 3 Phasen*. Zunächst sollen wenige Zielpersonen die beiden zu testenden Instrumente (Einführungsinterview und Haushaltsbuch) bearbeiten. Die durch fokussierte Intensivinterviews ermittelten Ergebnisse aus dieser Bearbeitung sollen gemeinsam mit den Ergebnissen aus den Fragebogenkonferenzen zu einer ersten Überarbeitung beider Instrumente führen. Die überarbeiteten Instrumente werden dann von erfahrenen Interviewern in 36 Haushalten mittels Beobachtungspretests, ergänzt durch Probing-Verfahren, überprüft.

2 Fokussierte Intensivinterviews

Die fokussierten Intensivinterviews wurden als unstandardisierte Befragungen auf der Basis allgemeiner Leitfragen offen durchgeführt. Wegen der Schwierigkeit einer schriftlichen Dokumentation des Gesprächs werden fokussierte Intensivinterviews bei ZUMA in der Regel per Tonbandmitschnitt aufgezeichnet, gelegentlich auch per Videokamera.

Das Design sah acht fokussierte Intensivinterviews vor. In jedem Haushalt kamen zwei Instrumente zum Einsatz, das Einführungsinterview und das Haushaltsbuch. Der Ablaufplan der Befragung sah vor, daß im ersten Schritt die Bezugsperson oder eine andere erwachsene Person des Haushalts das Einführungsinterview bearbeiten sollte. Im Anschluß daran sollte das fokussierte Intensivinterview über das Einführungsinterview durchgeführt und auf Tonband festgehalten werden. Danach sollte der Zielperson eine Einführung in das Ausfüllen des Haushaltsbuches gegeben werden. Das Haushaltsbuch sollte dann von der Zielperson über eine Zeitspanne von 5 bis 7 Tagen geführt werden. Beim Abholen gab es dann ein zweites fokussiertes Intensivinterview, ebenfalls per Tonband aufgezeichnet, dieses Mal bezogen auf das Haushaltsbuch. Um die zeit- und kostenintensive Abschrift der Bänder einzusparen, wurden die Bänder dann jeweils von einem Mitarbeiter der ZUMA-Feldabteilung gemeinsam mit einem Mitarbeiter des Statistischen Bundesamtes abgehört und interpretativ ausgewertet.

Die Zielhaushalte wurden durch ein einfaches Quotenverfahren ermittelt, wobei für die Auswahl der Ordnungsmerkmale die Erkenntnisse aus der EVS 1993 im Hinblick auf die Teilnahmebereitschaft der Haushalte herangezogen wurden. Im einzelnen wurden die Haushaltsgröße, die soziale Stellung und das Alter des Haupteinkommensbeziehers als Quotierungsmerkmale gewählt. Während Haushalte von Angestellten und Arbeitern bei der EVS 1993 eine schlechte Teilnahmebereitschaft aufwiesen, wurde erwartet, daß bei jungen Alleinerziehenden und Haushalten mit älteren Personen eher Schwierigkeiten beim Ausfüllen der Erhebungsunterlagen (zeitliche Probleme und Verständnisschwierigkeiten) auftreten würden. Der Haushaltstyp „Ehepaar mit 2 Kindern“ wurde dagegen wegen des erwarteten größeren Anschreibungsumfanges gewählt. Die Zielhaushalte sollten einem der folgenden Haushaltstypen zugeordnet werden:

- 1 - Personen-Haushalt, alleinerziehend, unter 30 Jahre, Arbeiter/-in.
- 1 - Personen-Haushalt, alleinerziehend, unter 30 Jahre, Angestellte/-r.
- Ehepaar über 60 Jahre, Haupteinkommensbezieher Arbeiter/-in.
- Ehepaar über 60 Jahre, Haupteinkommensbezieher Angestellte/-r.
- Ehepaar über 60 Jahre, Haupteinkommensbezieher Beamter/-in.
- Ehepaar mit 2 Kindern, Alter zwischen 30 und 55 Jahren, Haupteinkommensbezieher Arbeiter/-in.
- Ehepaar mit 2 Kindern, Alter zwischen 30 und 55 Jahren, Haupteinkommensbezieher Angestellte/-r.
- Ehepaar mit 2 Kindern, Alter zwischen 30 und 55 Jahren, Haupteinkommensbezieher Beamter/-in.

Außerdem wurde darauf geachtet, daß sich innerhalb der acht Haushalte sowohl Mieter wie auch Wohneigentumsbesitzer befanden, um Erkenntnisse über die Praktikabilität der Fragen zu den Wohnkosten zu erhalten.

Die eingesetzten Erhebungsunterlagen (Einführungsinterview und Haushaltsbuch) wurden vom Statistischen Bundesamt erstellt. Für die fokussierten Intensivinterviews entwickelte ZUMA geeignete „Leitfäden“.

Mit dem Einführungsinterview wurden für den Haushalt bzw. für jedes seiner Mitglieder Informationen aus verschiedenen Bereichen erhoben:

- Struktur des Haushalts (Alter, Geschlecht, etc.) jedes Haushaltsmitgliedes.
- Ausbildung (Schulabschluß, beruflicher Ausbildungsabschluß) jedes Haushaltsmitgliedes.
- Erwerbstätigkeit (Umfang, Art etc.) jedes Haushaltsmitgliedes.
- Sozialversicherung (Renten-, Kranken-, Pflegeversicherung) jedes Haushaltsmitgliedes.
- Unterhalt und Einkommen (Quelle, Höhe des Einkommens etc.) jedes Haushaltsmitgliedes.
- Wohnsituation (Wohnfläche, Art der Heizung, Kosten etc.) des Haushaltes.
- Vermögenssituation (Besitz von Vermögen, Kredite etc.) des Haushaltes.
- Ausstattung des Haushaltes mit Gebrauchsgütern.

Mit dem Haushaltsbuch sollten – ebenfalls teilweise für den Haushalt, teilweise für jedes seiner Mitglieder – folgende Bereiche erfaßt werden:

- Veränderungen in der Zusammensetzung des Haushaltes im Anschreibungszeitraum
- Einkommen aus unselbständiger Arbeit für jedes Haushaltsmitglied.
- Einkommen aus Rente/Pension für jedes Haushaltsmitglied.
- Übrige Einkommen/Einnahmen für jedes Haushaltsmitglied.
- Sachzugänge im Haushalt insgesamt.
- Detaillierte Erfassung aller Ausgaben des Haushalts pro Tag.
- Einnahmen und Ausgaben des Haushalts in den letzten beiden Monaten vor dem Anschreibungszeitraum.

Für jedes der beiden fokussierten Intensivinterviews wurde ein eigener Leitfragenkatalog entwickelt. Der Leitfaden für das Einführungsinterview beinhaltet unter anderem folgende Aspekte:

- Allgemeine Einstellungen zu Befragungen durch öffentliche Einrichtungen und Teilnahmebereitschaft.
- Äußerlichkeiten des Fragebogens wie Layout, Größe, Deckblatt.
- Verständlichkeit der Frageformulierungen.
- Verständlichkeit der Hinweise zu den Fragen.

- Allgemeine Einstellung zur Führung von Haushaltsbüchern.
- Allgemeine Informationen über die Abwicklung des Haushaltsalltags (Erledigung der Einkäufe und der finanziellen Angelegenheiten).

Im Leitfaden für das Haushaltsbuch wurde unter anderem auf folgende Bereiche abgezielt:

- Probleme beim Ausfüllen.
- Verständlichkeit der Fragen.
- Hilfsmittel zur Beantwortung der Fragen.
- Zeitbedarf.
- Allgemeine Verbesserungsvorschläge.
- Fragenabfolge.
- Teilnahmebereitschaft.
- PC-Besitz (im Hinblick auf die Führung des Haushaltsbuchs mit Hilfe des PC).

Diese Fragenbereiche wurden im Leitfaden noch weiter differenziert, wobei es der Methode entsprechend nicht erforderlich war, alle Fragen vollständig und akribisch abzuarbeiten. Vielmehr sollten Art, Inhalt und Abfolge der Frage sich dem Gesprächsablauf anpassen.

Die Feldarbeit wurde im April 1995 durchgeführt. ZUMA setzte insgesamt 5 Interviewer ein (vier wissenschaftliche Mitarbeiter der Feldabteilung und eine erfahrene Interviewerin des ZUMA-Interviewer-Stabes), die die Instrumente überbrachten und erläuterten, beim Bearbeiten des Einführungsinterviews anwesend waren und schließlich auch das fokussierte Intensivinterview zum Einführungsinterview durchführten. Die gleichen Interviewer holten nach der vereinbarten Ausfüllzeit das Haushaltsbuch beim Zielhaushalt wieder ab und führten das fokussierte Intensivinterview zum Haushaltsbuch durch.

Von den acht Zielhaushalten bearbeitete ein Haushalt das Haushaltsbuch 5 Tage lang, 4 Haushalte 6 Tage und 2 Haushalte 7 Tage lang. Der achte Haushalt brach nach 3 Tagen ab, weil der „Ausfüller“ erkrankte und sich niemand bereit erklärte, das Haushaltsbuch weiter auszufüllen. Das abschließende Intensivinterview konnte allerdings mit dem wieder genesenen „Ausfüller“ durchgeführt werden. Alle Haushalte erhielten nach Ableistung aller vereinbarten Arbeiten (Einführungsinterview, Haushaltsbuch, Teilnahme an den beiden fokussierten Intensivinterviews) ein Honorar von DM 50,-.

Zur Auswertung der Ergebnisse fanden insgesamt drei mehrstündige Sitzungen statt, bei der die eingesetzten Interviewer gemeinsam mit Vertretern des Statistischen Bundesamtes die Mitschnitte der fokussierten Intensivinterviews abhörten. Jeder Interviewer wertete die von ihm selbst geführten Interviews mit aus, um so gegebenenfalls noch Informationen an die Vertreter des Statistischen Bundesamtes weitergeben zu können, die den Tonbandaufnahmen nicht eindeutig zu entnehmen waren. Bereits aus dem gemeinsamen Abhören der Bänder und den in diesem Zusammenhang geführten Gesprächen ergab sich eine Vielzahl von Hinweisen zur Veränderung der eingesetzten Instrumente.

3 Fragebogenkonferenz

Das Instrument der Fragebogenkonferenz wurde bei diesem Pretest als Expertengespräch zur Beurteilung eines Fragebogens angewendet, wobei sich die Experten nicht nur „am grünen Tisch“ mit dem Fragebogen beschäftigten, sondern in aller Regel ein Eigeninterview sowie ein oder mehrere Pretestinterviews mit dem zu testenden Instrument durchführten.

Die Fragebogenkonferenz bei ZUMA fand an zwei Tagen im Mai 1995 statt. Der ursprünglich nicht vorgesehene zweite Tag mußte eingeschoben werden, weil es nicht gelungen war, beide Instrumente im Rahmen einer einzigen Sitzung zu bearbeiten. Teilnehmer der Fragebogenkonferenz waren die fünf Interviewer, die die fokussierten Intensivinterviews durchgeführt hatten sowie zwei weitere Interviewer, die die Bogen selbst testeten und zumindest ansatzweise Probeausfüllungen durch dritte Personen hatten durchführen lassen.

Im Statistischen Bundesamt führten die zuständigen Bearbeiter eine Fragebogenkonferenz durch, die zu zahlreichen Veränderungsvorschlägen und einer direkten Überarbeitung der Instrumente führte.

4 Ergebnisse der Intensivinterviews und der Fragebogenkonferenz

Insgesamt wurde eine Reihe von Veränderungsvorschlägen herausgearbeitet, von denen viele aus kleineren Anregungen zur Gestaltung der Erhebungspapiere bestanden. Als wichtigste Ergebnisse können festgehalten werden:¹⁾

Einführungsinterview

Der Fragebogen zum Einführungsinterview zeigte keine größeren Mängel, aber es gab zahlreiche kleinere Korrekturvorschläge. Daneben gab es folgende allgemeine Anmerkungen:

- Das Layout des Fragebogens ist zu vereinheitlichen. Ab Frage 17 sollte das doppelseitige Format der Fragen 1 bis 16 weitergeführt werden; außerdem sollten die Fragennummern für alle Fragen vom Layout her identisch sein.
- Alle Hinweise zur Erklärung der Fragen (linke Seite) sollten auf gleicher Höhe stehen wie die entsprechenden Fragen selbst (rechte Seite).
- Alle Überschriften im Fragebogen (wie z. B. „Ausbildung“, „Erwerbstätigkeit“ usw.) sind zu streichen.
- Bei allen Fragen mit mehr als zwei Antwortkategorien sollten eingefügt werden:
 - (Punkt) oder – (Strich).
- Die Filterhinweise (z. B. „Nur Mieter“ bei Frage 25 oder „Nur Eigentümer“ bei Frage 27) mußten generell deutlicher gesetzt werden (entweder mit folgender Leerzeile, größerer Schrift oder durch einen Rahmen).
- Alle linken Seiten sollten einen Hinweis enthalten, daß hier die Erklärung zu den einzelnen Fragen steht, alle rechten Seiten einen entsprechenden Hinweis darauf, daß hier vom Befragten etwas auszufüllen ist.

1) Um die Länge des Aufsatzes in einem angemessenen Rahmen zu belassen, können nur einige ausgewählte Ergebnisse berichtet werden.

- Das Einführungsinterview wurde in Form eines Selbstausfüller-Bogens konzipiert. Wünschenswert wäre es, wenn beim Einführungsinterview immer ein Interviewer anwesend sein könnte, um das Einführungsinterview und vor allem das Haushaltsbuch mit der Zielperson durchzusprechen und Erläuterungen zu geben. Wenn diese Möglichkeit besteht, sollte das Einführungsinterview als persönlich-mündliche Befragung geführt werden, nicht als Selbstausfüller.

Haushaltsbuch

Das Haushaltsbuch fand insgesamt eine positive Bewertung. Layout, Größe, Schrift und Format konnten beibehalten werden; allenfalls wurden größere Schrifttypen bei bestimmten Erklärungen (beispielsweise S. 38: „z. B. Inspektion“ bei „Reparaturkosten für Kraftwagen“) gewünscht.

Zu überlegen war, inwieweit die Sukzession des Haushaltsbuchs nicht nach folgendem Schema aufgebaut werden könnte:

Zunächst wird danach gefragt, ob es zwischenzeitlich Veränderungen im Haushalt gegeben hatte. Anschließend sollen die Einnahmen und Ausgaben der letzten zwei „befragungsfreien“ Monate eingetragen werden. Im Anschluß daran werden die Einkünfte erfaßt. Erst nach diesen Frageblöcken folgen die Seiten für die Sachzugänge und die im laufenden Monat getätigten Ausgaben.

Die Summe aller von ZUMA gemachten Vorschläge bildete gemeinsam mit den Überlegungen des Statistischen Bundesamtes die Grundlage für die Überarbeitung der Instrumente für Phase 3 des Pretest-Verfahrens, den Beobachtungspretest.

Die Ausfülldauer des Haushaltsbuches sollte einen Monat nicht überschreiten, da ansonsten nicht mehr in jedem Falle mit präzisen Angaben gerechnet werden könnte. Zu testen wäre, ob mit kleinerer Stichprobe und größeren finanziellen Incentives die „Präzision“ der Angaben zu steigern wäre.

Allgemeines

Den Befragten sollte eine Telefonnummer hinterlassen werden (Ansprechpartner bei den Statistischen Ämtern der Länder), um ihnen die Möglichkeit zu geben, bei Unklarheiten sofort Informationen abzurufen („Hotline“).

Bei der Umstellung des Verfahrens auf eine Zufallsstichprobe ist von einem telefonischen Screening-Verfahren abzuraten, da mehrere Zielpersonen ihre Bedenken gegen diese Form der Kontaktaufnahme bis hin zur potentiellen Verweigerungsbereitschaft geäußert hatten.

5 Der Beobachtungspretest

Nachdem das Einführungsinterview und das Haushaltsbuch aufgrund der Ergebnisse der fokussierten Intensivinterviews und der Fragebogenkonferenzen vom Statistischen Bundesamt überarbeitet worden waren, wurde ein Beobachtungspretest mit Probing-Verfahren (comprehension probes, information retrieval probes und confidence ratings) durchgeführt. Das Design des Beobachtungspretests sah vor, 36 Haushalte auszuwählen, die verschiedene Haushaltstypen repräsentieren sollten. In diesen 36 ausgewählten Haushalten sollte wieder mit dem Einführungsinterview begonnen werden, das nach wie vor als Selbstausfüller zu bearbeiten war. Von den Inhalten her unterschied sich das Einführungsinterview inhaltlich

nicht von dem in Phase 1 eingesetzten Fragebogen, doch waren unter technischen Aspekten und unter Gesichtspunkten der Handhabbarkeit des Instruments eine Vielzahl von Veränderungen vorgenommen worden.

Nachdem eine Haushaltsperson das Einführungsinterview bearbeitet hatte, sollte es zunächst zu einer Befragung der ausfüllenden Person zu dem gerade bearbeiteten Bogen kommen. Zu diesem Zweck entwickelte ZUMA einen Fragebogen, mit dem unter Anwendung unterschiedlicher Techniken Informationen zum Einführungsinterview eruiert werden sollten. Gefragt wurde nach der subjektiven Sicherheit der ausfüllenden Person bei der Angabe der geforderten Informationen für die einzelnen Haushaltsmitglieder (confidence rating), nach der Vollständigkeit der Angaben bei bestimmten Fragen des Einführungsinterviews (modifiziertes comprehension probe) und nach der Genauigkeit bestimmter Angaben. Schließlich sollte die ausfüllende Person Probleme mit den einzelnen Fragen angeben können, die nicht mittels dieser drei Verfahren bereits thematisiert worden waren.

Sieben Tage nach dem Einführungsinterview sollte der Interviewer die Zielperson erneut aufsuchen und sie mit Hilfe eines von ZUMA entwickelten Fragebogens zum Haushaltsbuch befragen. Dabei sollten ähnliche Techniken zum Einsatz kommen wie beim Fragebogen zum Einführungsinterview, diesmal nur bezogen auf das Haushaltsbuch und ergänzt um Fragen nach dem Erinnern von Käufen der vergangenen zwei Monate (information retrieval probes). Auch bei diesem zweiten Kontakt zwischen Interviewer und Haushalt sollte eine Haushaltsperson einen kurzen Selbstausfüller des Statistischen Bundesamtes bearbeiten, bei dem es um das Ausfüllen des Haushaltsbuches ging, um das Einkaufsverhalten und um allgemeine Fragen zur Führung eines Haushaltsbuchs.

Wieder sieben Tage später erfolgte der dritte Kontakt, der mit dem Abholen des Haushaltsbuches verbunden war. Dabei sollte die Vorgehensweise des zweiten Kontaktes repliziert werden. Durch den Einsatz des gleichen Fragebogens sieben und vierzehn Tage nach dem Einführungsinterview sollte festgestellt werden, ob sich die Einstellungen der ausfüllenden Haushaltsmitglieder bzw. der Ausfüller zu dem Haushaltsbuch und den damit verbundenen Anforderungen über die Zeit verändern würden.

Für den Beobachtungspretest wurden von 'ZUMA 36 Haushalte rekrutiert, wobei die Haushaltstypen zwischen ZUMA und dem Statistischen Bundesamt abgestimmt waren:

- 8 Alleinerziehende,
- 7 Ehepaare mit Kindern, Haupteinkommensbezieher bis 30 Jahre,
- 8 Ehepaare mit Kindern, Haupteinkommensbezieher zwischen 30 und 55 Jahren,
- 6 Allcinlebende ab 60 Jahre und
- 7 ältere Ehepaare, Haupteinkommensbezieher ab 60 Jahre.

Die Zielhaushalte sollten nach dieser allgemeinen Vorgabe der Auswahlkriterien von den ZUMA-Interviewern geworben werden.

Die Beobachtungspretests wurden im Mai/Juni 1995 im Großraum Mannheim durchgeführt. Zum Einsatz kamen 15 ZUMA-Interviewer. Alle Interviewer wurden in einer Schulung auf ihre Aufgabe bei dieser Befragung und die besonderen Schwierigkeiten dieser Studie vorbereitet.

Die Auswertung und Ergebnisermittlung erfolgte im wesentlichen bei ZUMA, nur die Auswertung der Haushaltsbücher selbst wurde direkt vom Statistischen Bundesamt übernommen, da nur dort geklärt werden konnte, ob die Eintragungen in ihrer Häufigkeit, vor allem aber in ihrer Differenziertheit und Genauigkeit den gewünschten Erfordernissen entsprachen.

Bei der Durchsicht der Einführungsinterviews zeigten sich nur ganz wenige, zumeist unbedeutende Probleme, z. B. vereinzelt Nichtbeantworten einer Frage oder eines Fragenteiles. Dies mag etwa daran gelegen haben, daß die Frage überlesen worden war oder daß jemand eine Antwort bewußt nicht geben wollte. Im nachhinein kann das nicht mehr geklärt werden und ist letztlich wegen des seltenen Vorkommens auch zu vernachlässigen. Alles in allem wurde das Einführungsinterview, sofern man dies aus der Durchsicht der Bogen ersehen konnte, formal und technisch weitgehend korrekt bearbeitet.

Einzig bei der Frage nach den individuellen Einkommen der Haushaltsmitglieder gab es eine explizite Verweigerung und in einem Fall wurde keine Angabe gemacht, weil die Information nicht vorlag. Häufiger leer blieb die dann folgende Frage nach dem gesamten Haushaltseinkommen: Von den 9 Fällen ohne Angabe bei dieser Frage hatten 6 Haushalte deshalb nichts mehr eingetragen, weil es sich um Einpersonenhaushalte oder um Alleinverdiener handelte, die geforderte Information somit schon bei der vorhergehenden Frage vollständig gegeben worden war. In einem Fall konnte das Haushaltseinkommen nicht angegeben werden, „da das Einkommen von einzelnen Haushaltsmitgliedern nicht bekannt ist“.

Insgesamt ist festzustellen, daß dieses Antwortverhalten als durchaus „diszipliniert“ bewertet werden kann, und daß die Häufigkeit des Nichtbeantwortens der Einkommensfrage eher im unteren Bereich dessen anzusiedeln ist, was aus anderen Befragungen bekannt ist. Die Notwendigkeit zu einer Modifikation der Einkommensfragen ergibt sich aus dem Pretest nicht.

Vereinzelt gab es Probleme mit der Differenzierung der Umlagen bei den Fragen zu Mieteinkünften, wobei diese Probleme aber sicherlich nicht durch fehlende Antwortbereitschaft entstanden sind, sondern durch fehlendes Wissen oder dadurch, daß eine Aufschlüsselung der Umlagen faktisch nicht existierte.

Ein technisches Problem ergab sich bei den Fragen, bei denen zeilenweise „Ja“ bzw. „Nein“ anzukreuzen war, die Befragten aber häufiger das „Nein“ ignorierten und die entsprechende Antwortkategorie leer ließen. Dies ist an sich nicht dramatisch, hätte aber dadurch eindeutiger werden können, daß man nur „Ja“-Antworten vorsah oder, falls das „Nein“ gewünscht war, explizit noch einmal darauf hätte hinweisen sollen, daß in jeder Zeile ein Kreuz gemacht werden sollte.

Ansonsten traten beim Einführungsinterview keine Probleme auf, wenn man einmal davon absieht, daß eine Person (Jahrgang 1915) beim Ausfüllen der Bogen aufgrund der kleinen Schrift Hilfe benötigte. Ergänzt man die Durchsicht der Einführungsinterviews um die Ergebnisse des Zusatzfragebogens von ZUMA, erweitert sich das Spektrum der Probleme etwas. Im folgenden werden einige Kommentare der Befragten in bezug auf die expliziten Nachfragen (ratings, probes) zu bestimmten Themenbereichen wiedergegeben.

Beruflicher Schulabschluß

Bei der Beantwortung dieser Frage fühlten sich einige Befragte unsicher. In einem Fall war unklar, worunter „Industriefachwirt“ (IHK-Prüfung) fällt, in einem anderen Fall das Studium an einer Pädagogischen Hochschule.

Soziale Stellung

Vier Befragte waren sich unsicher, wie die Frage zur sozialen Stellung zu beantworten war. In drei Fällen ging es um die Frage, wie Personen einzuordnen waren, die zum Zeitpunkt der Befragung in Erziehungsurlaub gewesen waren, also einen festen Arbeitsplatz hatten, aber aktuell nicht erwerbstätig waren.

Pflegeversicherung

Bei der Beantwortung dieser Frage fühlten sich 13 Befragte unsicher. Folgende Probleme traten auf:

- Ist die Pflegeversicherung Bestandteil der Krankenversicherung?
- Sind Rentner pflegeversichert?
- Was ist private Pflegeversicherung?
- Sind Kinder pflegeversichert?
- Sind Hausfrauen pflegeversichert?
- Ist man im Erziehungsurlaub pflegeversichert?

Die Frage nach der Pflegeversicherung war die einzige problematische Frage des gesamten Bogens, was aber nicht eine Folge der Operationalisierung oder der Technik war, sondern schlicht eine Folge der Tatsache, daß das Thema „Pflegeversicherung“ für die meisten Befragten (und nicht nur für diese) einen weißen Fleck auf der Wissenskarte darzustellen schien. Für die Frage zur Pflegeversicherung bedurfte es einer einfachen Erklärung, was eigentlich die Pflegeversicherung genau sei. Die Erklärung sollte einen Hinweis darauf enthalten, wie die „Problemgruppen“ Kinder, Hausfrauen, Rentner, Pensionäre pflegeversichert sind.

Nettoeinkommen

Einige unregelmäßige Einkommenszahlungen wurden von 15 Befragten nicht berücksichtigt:

- Weihnachtsgeld und Urlaubsgeld, weil es sich dabei um einmalige Zahlungen handelte, die zum Zeitpunkt der Befragung noch nicht feststanden oder weil es nichts mit dem monatlichen Nettoeinkommen zu tun hatte (zwei Nennungen) oder weil es einfach vergessen wurde (zwei Nennungen) oder weil gar nicht bekannt war, was der neue Arbeitgeber zahlen würde.
- Kindergeld, weil es einfach vergessen wurde (zwei Nennungen) oder weil es nichts mit dem normalen monatlichen Nettoeinkommen zu tun hatte oder weil der Fragebogen nicht ordentlich durchgelesen wurde.

Mietkosten

Von den 19 Mietern der Befragung konnten 6 die Mietkosten nicht „ganz exakt“ angeben. Gründe hierfür waren:

- Unklarheiten mit den Nebenkosten, Unterlagen waren beim Steuerberater.
- Bei Umlagen keine genaue Aufschlüsselung möglich (drei Nennungen), Angaben grob geschätzt.
- Bestimmte Umlagen wurden unregelmäßig erhoben und waren daher nicht auf den Monat umrechenbar gewesen.

Kosten Eigentumswohnung

Von den 6 Besitzern von Eigentumswohnungen konnten 2 die Kosten nicht „ganz exakt“ angeben, einer davon, weil die Umlagen alle acht Wochen vollständig bezahlt und dann auf alle im Haus lebenden Parteien umgelegt worden waren.

Zusatzbefragung des Statistischen Bundesamtes

Aus der Zusatzbefragung des Statistischen Bundesamtes zum Einführungsinterview sollten unter anderem erste Erkenntnisse über die Motivation der Haushalte, die Bereitschaft zur Nutzung von DV-gestützten Erhebungsinstrumenten und die Führung von Haushaltsbüchern gewonnen werden. Im folgenden sollen nur einige ausgewählte Ergebnisse dargestellt werden. Nur in einem der 36 Haushalte wurde am Tage des Einführungsinterviews bereits ein privates Haushaltsbuch geführt, 11 Haushalte gaben an, sie hätten früher hin und wieder schon einmal ein Haushaltsbuch geführt, 22 Haushalte hatten noch nie ein Haushaltsbuch geführt, zwei weitere wollten dies schon immer einmal tun, hatten diesen Plan aber nie umgesetzt.

Von den 17 Haushalten, die über einen eigenen PC verfügten, waren nur 5 bereit, die Eintragungen mit dem PC zu tun; 9 der 12 anderen Haushalte begründeten ihre mangelnde Bereitschaft dazu damit, daß dies „zu aufwendig“ sei. Dies läßt allerdings vermuten, daß der Hinweis „mit Hilfe eines benutzerfreundlichen Programms“ im Fragetext entweder nicht richtig beachtet worden war oder daß die Befragten keine Vorstellung darüber hatten, wie benutzerfreundlich und damit aufwandsmindernd ein solches Vorgehen sein würde. Ein Haushalt lehnte die PC-Option ab, weil die Vergütung zu gering war.

Warum haben die Haushalte an der Befragung teilgenommen? Die meisten der (Mehrfach-) Nennungen bezogen sich auf das Interesse an der Studie (20 Nennungen), knapp gefolgt von dem finanziellen Aspekt, sich etwas hinzuverdienen zu wollen (18), der generellen Teilnahmebereitschaft an Befragungen (15) und schließlich dem Bemühen, sich durch die Teilnahme an unserer Befragung einen Überblick über die eigene wirtschaftliche Situation zu verschaffen (12).

Beantwortet wurden die Fragen des Einführungsinterviews zur Hälfte vom Haupteinkommensbezieher des Haushaltes, zur Hälfte vom Ehepartner. Andere Personen spielten keine Rolle.

Die meisten Schwierigkeiten beim Ausfüllen gab es bei den Fragen zur Pflegeversicherung (13 Nennungen); hier wurden Unterlagen (3) und weitere Personen (6) zu Rate gezogen.

Probleme, aber deutlich weniger als bei der Pflegeversicherung, gab es auch bei den Fragen zur Miete (8), zum Einkommen (6), zur Rentenversicherung und zum beruflichen Ausbildungsabschluß (4), zum Schulabschluß und zur Krankenversicherung (3).

Das Haushaltsbuch

Die inhaltliche Durchsicht und Auswertung der Angaben in den Haushaltsbüchern erfolgte beim Statistischen Bundesamt. Hierbei stellte sich heraus, daß einigen Befragten die Tiefe der Anschreibungen unklar war. Künftig wären daher zusätzliche Informationen im Haushaltsbuch oder in einer Beilage hilfreich. Die Auswertung der Zusatzfragebogen (für 7 und 14 Tage) wurde bei ZUMA durchgeführt. Im folgenden werden einige allgemeine Hinweise, die sich aus den Kommentaren der Befragten und der Durchsicht der Haushaltsbücher ergeben haben, dargestellt:

Zusatzbefragung zum Haushaltsbuch nach 7 Tagen

Farbe, Größe der Schrift

- Schriftgröße sollte verändert werden.
- Farbliche Differenzierung zwischen Sachzugängen und Ausgaben wegen Verwechslungsgefahr.
- Kleineres Format wünschenswert.

Inhalte

- Nicht zuordenbare Kosten, Urlaubskosten bzw. Hotelkosten waren nicht zu differenzieren.
- Filter einbauen, beispielsweise bei der Art des Einkommens, so daß es eindeutig möglich ist, nicht zutreffende Einkommensarten (Rente/Pension o. ä.) zu überspringen.

Weitere Vorschläge

- Man könnte vielleicht innerhalb der Ausgabenspalte Unterteilungen machen für Lebensmittel, Reinigungsmittel, Lesestoff etc., dann würde man auch bestimmt nicht so leicht etwas vergessen.
- Häufiger Definitionen zu bestimmten Einkommensarten oder Begriffen (Termingeldkonto, Deputat etc.) geben.
- Buch könnte kleiner sein (DIN A 5-Format).
- Stabilere Papierzusammenhaltung wäre sehr gut.

Zusatzbefragung zum Haushaltsbuch nach 14 Tagen

Farbe, Größe der Schrift

- Sachzugänge und tägliche Eintragungen farblich unterschiedlich gestalten, täglich einen Bereich „Besonderheiten“ für außergewöhnliche Dinge wie „Sachzugänge höher an Geburtstagen“ usw. hinzufügen.
- Das Haushaltsbuch sollte handlicher sein.

Inhalte

- Hinweise noch deutlicher und mehr Definitionen zu bestimmten Begriffen nennen.
- Die Angaben nach Rubriken (Lebensmittel, Kraftstoff, sonstiges) unterteilen, dies sei einfacher für den Ausfüller oder dies gäbe die Möglichkeit, die Kassenbons auszufüllen.

Die Angaben aus dem Probing-Verfahren sind im Anhang dargestellt. Dabei wird nach den Ergebnissen, die sich aus dem nach 7 und nach 14 Tagen eingesetzten Verfahren ergeben haben, unterschieden.

Es zeigt sich ein Gewöhnungseffekt beim Ausfüllen des Haushaltsbuches. 11 der ausfüllenden Personen gaben bei der Endbefragung an, das Führen des Haushaltsbuches sei ihnen gegen Ende der Anschreibungszeit leichter gefallen als nach 7 Tagen. Für 23 Haushalte gab es keinen Unterschied, nur in einem einzigen Fall wurde berichtet, daß das Ausfüllen mit der Zeit schwerer gefallen sei.

Die Auswertung des Zusatzfragebogens des Statistischen Bundesamtes führte zu weiteren Erkenntnissen über das Haushaltsbuch. So wurde das Haushaltsbuch während des gesamten Anschreibungszeitraums immer von der gleichen Person ausgefüllt, zumeist (nach 7 Tagen: 31 Nennungen, nach 14 Tagen: 30) von der Ehe-/Frau. Nach 14 Tagen wurde für das Ausfüllen des Haushaltsbuches weniger Zeit beansprucht als nach 7 Tagen, sowohl bei der Erfassung der kleineren Einkäufe als auch – etwas weniger deutlich – bei der Erfassung der Großeinkäufe:

**Tabelle 1: Dauer des Ausfüllens im Haushaltsbuch
bei kleineren Einkäufen und bei Großeinkäufen**

Ausfülldauer von ... bis unter ... Minuten	Kleinere Einkäufe		Großeinkäufe	
	nach 7 Tagen	nach 14 Tagen	nach 7 Tagen	nach 14 Tagen
Anzahl der Nennungen				
bis 5	18	22	2	1
5 bis 10	16	11	4	9
10 bis 15	2	1	12	7
15 bis 20			7	4
20 bis 30			1	2
30 bis 45				1
45 und mehr			1	1

Kleinere Einkäufe wurden nach 7 Tagen eher unmittelbar nach dem Einkauf registriert, nach 14 Tagen eher zu einem unbestimmten Zeitpunkt am Tage des Einkaufes. Nur in vereinzelten Fällen vergingen 1 bis 2 Tage, bis der Einkauf eingetragen wurde.

Schwierigkeiten beim Bearbeiten des Haushaltsbuches ergaben sich nach 7 Tagen vor allem bei den Angaben zu den beiden vorangegangenen Monaten, beim Einkommen und bei den Sachzugängen; in großem Maße wurde auf Unterlagen oder die Hilfe weiterer Personen zu-

rückgegriffen. Die Schwierigkeiten beim Ausfüllen gingen nach 14 Tagen aber deutlich zurück; auch die zusätzlichen Informationsquellen wurden nicht mehr ganz so oft beansprucht. Dies erklärt sich sicherlich vor allem dadurch, daß die Angaben zu den beiden zurückliegenden Monaten, die doch deutlich schwieriger waren als das tägliche Registrieren der Ausgaben, gleich zu Beginn der Feldarbeitszeit gemacht wurden, sich deshalb auch nur bei der Befragung nach 7 Tagen auswirkten, aber nicht mehr bei der Endbefragung.

Tabelle 2: Probleme beim Ausfüllen des Haushaltsbuchs im Zeitablauf *)

Thematik im Haushaltsbuch	Schwierigkeiten bei Fragestellungen		Inanspruchnahme von Unterlagen beim Aus- füllen		Hinzuziehen von an- deren Personen beim Ausfüllen	
	7 Tage	14 Tage	7 Tage	14 Tage	7 Tage	14 Tage
	Anzahl der Nennungen					
Angaben zu beiden vor- angegangenen Monaten	6	3	10	10	7	5
Einkommen	5	2	9	8	7	5
Rente/Pension	4	1	6	3	7	5
Öffentliche Zahlungen	1	1	8	7	2	0
Weitere Einnahmen	1	0	2	0	4	3
Sachzugänge	6	1	0	1	2	1
Ausgaben	1	0	7	6	3	1
Keine Schwierigkei- ten/keine Nennungen .	7	14			6	5

*) Jeweils 7 bzw. 14 Tage nach Beginn des Führens des Haushaltsbuchs.

Mit 24 Ja-Stimmen war nach 7 Tagen die Bereitschaft erfreulich hoch, erneut an der Befragung teilzunehmen. Diese Bereitschaft änderte sich auch nach 14 Tagen nicht. Dominantes Argument gegen eine weitere Teilnahme war „kein weiteres Interesse“ mit zunächst 7, dann 9 Nennungen. Die Vergütung wurde nur als Randproblem gesehen, zunächst drei, dann noch zwei Haushalte hielten sie für zu gering.

Das Haushaltsbuch in der vorliegenden Form erfüllt weitgehend seine Funktion. Als vorteilhaft hat sich die Veränderung der Sukzession gegenüber der ersten Version erwiesen (also zuerst die beiden vorangegangenen Monate abarbeiten zu lassen und erst dann den Anschreibungsmonat zu beginnen).

Der Beobachtungspretest ergab eine größere Zahl an Anmerkungen und Hinweisen auf kleinere Probleme, die aber nicht die Notwendigkeit ergaben, das Haushaltsbuch insgesamt noch einmal zu modifizieren. Man muß vielmehr einfach zur Kenntnis nehmen, daß mit einem Haushaltsbuch grundsätzlich nicht jegliche Facette der komplexen Information über die gesamten finanziellen Bewegungen jedes einzelnen Haushaltes erfaßt und detailliert jede denkbare Konstellation beschrieben werden kann.

6 Fazit

Die getesteten Instrumente für die Neukonzeption der laufenden Wirtschaftsrechnungen können nach Abschluß des Pretests als weitgehend feldreif angesehen werden; dies gilt sowohl für das Einführungsinterview als auch für das Haushaltsbuch. Vereinzelte Probleme, bei denen noch etwas nachzubessern wäre, sind deutlich geworden (z. B. Pflegeversicherung). Eine größere Zahl kleiner und kleinster Kommentare kann zur Kenntnis genommen und abgehakt werden, weil es nicht möglich ist, ein Befragungsinstrument bis ins kleinste Detail auf alle denkbaren Situationen auszurichten, noch dazu bei einem so komplexen Sachverhalt wie dem hier zur Diskussion stehenden.

Als großer Schritt nach vorne muß ganz sicherlich die Überarbeitung der Instrumente aufgrund der Ergebnisse der fokussierten Intensivinterviews bewertet werden; der Beobachtungspretest zeigte, daß dadurch ein weitgehend feldreifes Instrument entstanden ist, das ohne allzuvielen größeren Schwierigkeiten in 36 Fällen zufriedenstellende Ergebnisse brachte. Dieses Ergebnis spricht für das praktizierte „multi method pretesting“, also die Kombination unterschiedlicher Pretestverfahren.

Anhang

Dokumentation der Angaben aus den „Probing-Verfahren“ *)

Probleme mit den Einnahmen der letzten beiden Monate

Nach 7 Tagen:

- Keine Probleme (33).
- Zinseinkommen von Sparbüchern war nicht bekannt und kein Interesse, dies extra auszurechnen oder nachzuschauen (1).
- Bei jährlichen Auszahlungen: Sparbuch wird wie Girokonto benutzt (1).
- Die genaue Auflistung der Abzüge vom Einkommen (1).

Nach 14 Tagen:

- Keine Probleme (35).
- Kurpauschale einmalig vom Sozialamt erhalten. Weiß nicht, wo ich das eintragen sollte (1).

Vorgehensweise beim Erinnern der Einnahmen der letzten beiden Monate

Nach 7 Tagen:

- Gemeinsam mit Ehepartner u.ä. (3).
- Die Zeit vor dem geistigen Auge Revue passieren lassen (1).
- Kontoauszüge, Unterlagen (15).
- Sozialhilfebescheid (2).
- Immer gleiche Einnahmen (4).
- Wissen (1).
- Nicht zuordenbare Angaben (3).
- Keine Angaben (2).
- Nachdenken/Erinnerung (13).
- Schätzung (1).

Nach 14 Tagen:

- Keine wesentlichen Änderungen.

Probleme mit den Ausgaben der letzten beiden Monate

Nach 7 Tagen:

- Keine Probleme (27).
- Die entsprechenden Unterlagen mußten kurz herangezogen werden (1).
- Zum Punkt „Ausgaben für Energie“: Elektrizität und Gas konnte nicht getrennt werden. Zum Punkt „Umlagen“: Treppenhausbeleuchtung ist bei der Miete enthalten und nicht bei der Stromrechnung (2).
- Worunter fällt die Grundsteuer? Essengehen? (1)

*) In Klammern jeweils Anzahl der Nennungen.

- Bildung, Unterhaltung, Freizeit: Ist hier Fachliteratur eingeschlossen, Software, Hardware? (1)
- Strom und Gas mußte auf der Rechnung nachgeschaut werden (1).
- Probleme bei den Kreditzahlungen, weil das immer der Mann macht (1).
- Gas und Elektrizität wird als Pauschale bezahlt (1).

Nach 14 Tagen:

- Keine Probleme (30).
- Probleme wie nach 7 Tagen (2).
- Punkt Energie: Keine Aufschlüsselung für Elektrizität und Gas möglich. Die Umlage für Hausbeleuchtung konnte nicht näher differenziert werden, da in Miete pauschal enthalten (1).
- Grundsteuer bei sonstigen Ausgaben, da bei „Steuern“ die Möglichkeit „Grundsteuer“ fehlt (1).
- Hobbyartikel (Aquarium) waren nicht eindeutig zuzuordnen (1).

Ungenauere Angaben zu Einkommen aus selbständiger Arbeit

Nach 7 und nach 14 Tagen:

- Keine exakten Angaben möglich (4).

Probleme mit den täglichen Ausgaben

Nach 7 Tagen:

- Keine Probleme (21).
- Teilweise war der Kassenschein nicht mehr da, so daß genaue Angaben nicht mehr gemacht werden konnten (1).
- Immer rechtzeitig daran zu denken, daß die Eintragungen erfolgt sind, ohne daß man das ganze Zeug schon weggestellt hat und den Bon weggeworfen hat (1).
- Kassenzettelausgabe gibt es nicht in allen Läden bzw. Märkten. Dadurch sind manchmal die Artikel zusammengefaßt mit einem Preis ausgegeben. Die Auseinandergliederung in Einzelausgaben war im nachhinein nicht mehr möglich (1).
- Disziplin für tägliche Eintragungen zu halten (1).
- Problem der genauen Aufsplittung in Einzelwaren und Zuordnung der Preise (1).
- Beim Einkaufen auf dem Markt gab es keine Auflistung der einzelnen Produkte. Habe dann die Preise geschätzt (1).
- Nicht ganz klar, wie genau die Angaben sein sollten: Kartoffeln oder Speisefrühskartoffeln, mit oder ohne Gewichtsangabe? (1)
- Überweisungen und Daueraufträge wurden fast vergessen (1).
- Einzelne Auflistungen oder Zusammenfassungen? (1)
- Umständlich, viel Aufwand, alles einzutragen, so im Detail (1).
- Bei einem Kassenschein, wo nur die Preise draufstehen, fällt es etwas schwer, die einzelnen Posten, die einzelnen Produkte wiederzuerkennen (1).

Nach 14 Tagen:

- Keine Probleme (24).
- Bei Einkäufen, bei denen es keine Kassenzettel gab, die Gesamtsumme nach Einzelpreisen auszdifferenzieren (1).
- Wußte nicht genau, ob Oberbegriffe „Metzger“, „Bäcker“ reichen, oder ob ich das genau auflisten soll (1).
- Die laufenden Zahlungen wie Miete, Strom ... sind Daueraufträge, die hier einzutragen vergessen worden sind, da diese automatisch abgebucht werden, und man nicht ständig Kontoauszüge holt (1).
- Disziplin für tägliche Eintragungen halten (1).
- Aufsplitten der einzelnen Waren, wenn Kassenbon nicht explizit alles auflistet (1).
- Wenn auf Rechnung keine Auflistung der einzelnen Produkte, dann habe ich geschätzt (1).
- Falls bei Kassenbon keine Artikelbeschreibung ausgedruckt war, Zuordnung schwierig (1).
- Manchmal nicht ganz klar, wie genau die Angaben sein sollten (1).
- An Daueraufträge und Überweisungen zu denken, ob und wie Pfandgeld aufgeschrieben werden soll, ob und wie ich Vorlage und Erstattungsbeträge eintragen soll (1).
- Bei Eintragungen, die nicht selbst vorgenommen worden sind, fällt Erinnerung – wenn kein Kassenbon vorhanden – schwer (1).
- Zu aufwendig, es wäre einfacher, wenn es schon Rubriken gäbe (Milch, Wurst, Fleisch, Lebensmittel ...), habe extra wenig eingekauft (1).

Sonstige Probleme mit dem Ausfüllen

Nach 7 Tagen:

- Keine Probleme (27).
- Interviewer: Bei den Eintragungen fehlt es deutlich an der Lust. Vermutlich war es zu anstrengend, sich die Artikel auf dem Kassenbon genau rauszusuchen und einzutragen (1).
- Geburtstag eines Familienmitgliedes: Frau kauft, Mann bekommt Geschenke (Sachzugänge?) (1).
- Umlage der Kosten, die jährlich anfallen, wie z.B. Versicherungen (1).
- Ausgaben, die bei der täglichen Arbeit bzw. beim Studium anfallen, sich zu merken bzw. Preise einzuordnen (1).
- Habe ein Wochenende bei Freunden zugebracht und somit für Kost und Unterkunft keine Ausgaben. Habe dies allerdings auch nicht bei Sachzugängen (3 Mahlzeiten) aufgelistet (2).
- Habe DM 500,- von der Bank erhalten (Anteile eines von mir aufgelösten Kontos), wußte nicht, wo ich das vermerken sollte. Habe von Krankenkasse Kursgeld rückerstattet bekommen, habe ich bisher nicht aufgelistet (1).
- Bei Sachzugängen: Tausch – Kennbuchstabe fehlt. Sachzugänge: Da Großtierhaltung und Schlachtung der Rinder beim Metzger (Teil des Fleisches erhält der Züchter zurück) und deshalb Fleisch, das verbraucht wird aus diesem Geschäft, konnte nicht eingeordnet werden unter den vorgegebenen Punkten (1).
- Mir ist unklar, wie ich mit den Tagen umgehen soll, an denen ich nichts ausgabe (1).

Nach 14 Tagen:

- Keine Probleme (34).
- Befragte hat die Tage, an denen sie keine Ausgaben hatte (Sonn- und Feiertage), nicht eingetragen. Es sind zwar nur 11 Tage eingetragen, das Haushaltsbuch wurde aber insgesamt über 14 Tage geführt (1).
- Unklarheiten über die Genauigkeit der Eintragungen (Mengenangaben, Flohmarktartikel u. ä.) (1).

Fragebogen-Pretest zur „Erfassung von Produktionstagen und Betriebszeiten“

Vorbemerkung

Die Verkürzung der Arbeitszeit, die zunehmende Bedeutung der Teilzeitarbeit und die verstärkte Arbeitszeitflexibilisierung haben eine Entkoppelung von individueller Arbeitszeit der Beschäftigten und Betriebszeit der Produktionsanlagen zur Folge, da viele Betriebe ihre Anlagen auf Mehrschichtsysteme umstellen. Die ökonomischen Vorteile einer Verlängerung der Betriebszeiten liegen unter anderem in einer Kostendegression und in der Erhöhung der Produktivität und der damit verbundenen Steigerung der Wettbewerbsfähigkeit gegenüber anderen Produzenten.

Über die reale Veränderung der Betriebszeiten wird viel spekuliert, ohne sich jedoch auf gesicherte Daten stützen zu können, denn die Produktionsstatistiken des Statistischen Bundesamtes stellen bisher keine regelmäßigen Angaben über Produktionszeiten bereit. Informationen über die Entwicklung der effektiven Produktionszeiten sind jedoch insbesondere im Hinblick auf die Entwicklung eines Europäischen Binnenmarktes von großer Bedeutung. Es ist bisher kaum möglich, internationale Vergleiche von Produktionszeiten und damit von Produktivitätsgrößen durchzuführen.

Die unterschiedliche Behandlung von Ferien- und Brückentagen führt schon bei der Produktionszeit zu Erfassungsproblemen. Eine Kalenderbereinigung, die lediglich die Anzahl der Werktage nach dem Kalender betrachtet, reicht hier nicht aus. Zudem wird die effektive Produktionszeit von der Nutzung verschiedener Schichtarbeitstypen beeinflusst. Die daraus resultierenden Schwierigkeiten bei der Berechnung – die Nutzungsintensität eines Arbeitsplatzes müßte einbezogen werden – machen die Konstruktion von Hilfsmerkmalen erforderlich. Diese sind die Anzahl der Produktionstage und die Betriebszeit der Produktionsanlagen. Bei Betrieben mit unterschiedlicher Schichtbesetzung ist zudem eine Gewichtung der Betriebszeiten mit der Zahl der jeweils Beschäftigten vorzunehmen, um aussagefähige Daten zu erhalten.

Vom Statistischen Bundesamt wurden drei Erhebungsbogen entwickelt, die die Merkmale „Produktionstage im Berichtsmonat“ und „Betriebszeit der Produktionsanlagen im Berichtsmonat“ erfragen sollten. Für den Pretest wurden die folgenden Ziele vorgegeben:

- Alle Erhebungsbogen sollten in unterschiedlichen „Branchen“ getestet werden.
- Die Akzeptanz der jeweiligen Fragestellung und des gesamten Bogens sollte ermittelt werden.

^{*)} Dr. Manfred Ehling, Statistisches Bundesamt, Wiesbaden.
Rolf Porst, Zentrum für Umfragen, Methoden und Analysen (ZUMA), Mannheim.
Elmar Wein, Statistisches Bundesamt, Wiesbaden.

- Aussagen über die Schwierigkeiten bei der Beantwortung jeder einzelnen Frage (z. B. Probleme bei der Datenerfassung, Aufwand der Datenbeschaffung im Unternehmen) sollten gemacht werden können.
- Verbesserungsvorschläge hinsichtlich der Methodik und der Gestaltung des Erhebungsbogens sollten von Unternehmen aus unterschiedlichen Branchen und verschiedenen Beschäftigtengrößenklassen vorliegen.

Am Ende des Pretests sollte eine konkrete Empfehlung zur Entscheidung für eines der drei zu testenden Instrumente (Erhebungsbogen) bzw. eine konkrete Empfehlung für ein neu zu erstellendes Instrument gegeben werden.

1 Konzeption der Studie

Im Rahmen des Pretests sollten drei unterschiedliche Versionen des Erhebungsbogens für die „Erfassung von Produktionstagen und Betriebszeiten“ getestet werden. Mit jedem Erhebungsbogen sollten 20 Befragungen durchgeführt werden, insgesamt also 60. Zu befragen waren Inhaber und Geschäftsführer (bzw. vergleichbare Personen) von Unternehmen.

Die Zielpersonen sollten zunächst durch ein telefonisches Screening ermittelt werden. Dabei sollte zugleich ein Termin für die Übergabe des Erhebungsbogens vereinbart werden, der aber nicht in jedem Falle zugleich auch der „Befragungstermin“ sein würde. Falls die Zielperson zum Ausfüllen viel Zeit benötigte, auf Unterlagen zurückgreifen oder nachschlagen mußte, hätte der Interviewer den Erhebungsbogen zum Selbstaussüllen vor Ort hinterlassen und einen Termin vereinbart, zu dem er den Bogen wieder abholen und zugleich die Zielperson zu dem Bogen persönlich mündlich befragen würde. Im Gegensatz zu den üblichen Beobachtungspretests wäre der Interviewer während des Ausfüllens durch die Zielperson nicht anwesend. Damit ginge zwar ein wichtiger Teil der Beobachtungsfunktion der Interviewer verloren, durch die Nachbefragung dürften jedoch ausreichend auswertungsrelevante Hinweise erzielt werden.

2 Stichprobenverfahren

Die Grundgesamtheit der zu befragenden Betriebe wurde über ihre Branchenzugehörigkeit und über die Anzahl der Beschäftigten definiert; einzubeziehen waren Betriebe mit 100 oder mehr Mitarbeitern, da ab dieser Beschäftigtenzahl häufiger mehrere Schichten je Betrieb erwartet wurden. Darüber hinaus sollten die Betriebe folgenden Zweistellern der Systematik der Wirtschaftszweige, Fassung für die Statistik im Produzierenden Gewerbe (SYPRO), zugeordnet werden können:

SYPRO-Zweisteller	Bezeichnung
25	Gewinnung und Verarbeitung von Steinen und Erden;
32	Maschinenbau;
36	Elektrotechnik, Reparatur von elektrischen Geräten für den Haushalt;
40	Chemische Industrie;
68	Ernährungsgewerbe.

Innerhalb der Betriebe sollten Personen befragt werden, die kompetent über Produktionstage und Betriebszeiten Auskunft würden geben können. Die auszuwählenden Betriebe sollten, um die Kosten für die Feldarbeit akzeptabel zu halten, auf ausgewählte Postleitzahlenbezirke begrenzt werden. Da es erwartungsgemäß sehr schwierig sein würde, die Interviews bei dieser Spezialpopulation zu realisieren, und da für einen Instrumententest dieser Art die Frage nach Stichprobe und Ausschöpfung ohnehin nur von nachrangigem Interesse war, sollte eine entsprechend hohe Zahl von Firmen, deren Adressen ermittelt und deren Zugehörigkeit zur Grundgesamtheit mit Bestimmtheit festgestellt werden konnte, prinzipiell als zur Befragung zur Verfügung stehend definiert werden und damit in die Stichprobe eingehen.

Für die Stichprobenziehung kamen schließlich Adressen zum Einsatz, die von einem Adreßverlag nach den Merkmalen räumliche Zuordnung, Branchenzuordnung und Beschäftigtenzahl ermittelt und zur Verfügung gestellt wurden. Alles in allem lagen zunächst 407 Firmenadressen vor, wobei insbesondere die Frage, ob deren Beschäftigtenzahl die 100-Beschäftigten-Grenze überschreiten würde, durch Sichtung des Adreßmaterials nicht immer unbedingt eindeutig und unmittelbar zu bejahen war. Deshalb wurde das Kriterium Beschäftigtenzahl zu einem der zentralen Fakten in der Phase der Kontaktierung der Firmen durch die Interviewer (tatsächlich zeigte sich dann auch, daß 29 Firmen weniger als 100 respektive 75 Beschäftigte (bei Firmen des SYPRO-Zweistellers 25) hatten und deshalb aus der Stichprobe zu nehmen waren).

3 Erhebungs- und Befragungsinstrumente

Die vom Statistischen Bundesamt vorgelegten zu testenden drei Versionen des Erhebungsbogens wurden zunächst vom Zentrum für Umfragen, Methoden und Analysen (ZUMA) kommentiert; diese Kommentare und Vorschläge sind bei einer anschließenden Überarbeitung berücksichtigt worden. Die drei Versionen des Erhebungsbogens sind auf den folgenden Seiten (siehe S. 94 ff.) abgedruckt.

Neben den eigentlichen Erhebungsbogen mußte ein Fragebogen entwickelt werden, mit dem die Interviewer die Zielpersonen zum Erhebungsbogen befragen sollten. Hierzu wurde ein Konzept durch das Statistische Bundesamt vorgelegt, von ZUMA überarbeitet, in Fragebogenformat gebracht und gestaltet. In einer gemeinsamen Fragebogenendredaktion wurde dieser Fragebogen dann verabschiedet.

Von ZUMA wurde schließlich das Kontaktprotokoll erstellt sowie ein Blatt, auf dem die Interviewer ihre Eindrücke von der Befragungssituation dezidiert schildern sollten („Situationsbogen“). Als Argumentationshilfe für die Interviewer und zur Vorlage bei den Zielpersonen gab es schließlich noch ein kurzes Schreiben, in dem der Zweck der Studie dargestellt, die Anonymität zugesichert und Rückruf-Telefonnummern angegeben waren.

Erhebungsbogen Version A – Vorderseite

Statistisches Landesamt	Erfassung von Produktionstagen und Betriebszeiten im Verarbeitenden Gewerbe	
	Betriebs-Nr.: Bei Rückfragen bitte angeben.	
Unsere Telefon-Nr.:	Bitte senden Sie 1 Exemplar bis zum 5. des folgenden Monats zurück.	
Name und Anschrift des Betriebes:	Rechtsgrundlagen, Erläuterungen und Beispiele finden Sie im Anschluß an die Fragen.	
	Name eines Ansprechpartners:	
	Telefon-Nr.:	
Betriebs-Nr. (SST 1 - 9):	SYPRO-Nr.:	Berichtsmonat:

Beachten Sie bitte die Erläuterungen und Beispiele auf der Rückseite!

1. Berechnung der Produktionstage

Für die Berechnung dieses Merkmals werden die Tage eines Berichtsmonats, an denen keine oder nur eine reduzierte Produktion stattfand, benötigt.

Es fand an Tagen des Berichtsmonats keine Produktion statt.

2. Berechnung der Betriebszeiten

Mit der Betriebszeit ist die monatliche Stundenzahl gemeint, in der in den Produktionsbereichen Ihres Betriebes gearbeitet worden ist. Zu diesen Stunden zählen auch Vorbereitungs- und Hilfszeiten.

Die Betriebszeit der Produktionsbereiche beträgt im Berichtsmonat:

SYPRO-Nr. des Betriebsteils	Monatliche Betriebszeit in Stunden	Personal	Arbeitsplätze

Erhebungsbogen Version A – Rückseite

Erläuterungen und Beispiele

1. Produktionstage

Der Betrieb XYZ produziert in seinem fachlichen Betriebsteil "Produktion" täglich in 8 Stunden mit 200 Arbeitnehmern.

An einem Monat mit 31 Kalendertagen wurde an 3 Samstagen und 4 Sonntagen nicht produziert. Es entstanden damit 7 fehlende Tage.

An einem Freitag arbeiteten nur 100 Arbeitnehmer. Der Freitag ist daher als ein halber Fehltag zu werten.

An einem Dienstag wurde nur über 4 Stunden produziert. Der Ausfall von 4 Stunden ist daher als ein halber Fehltag zu werten.

Im Betrieb XYZ fand damit an 8 ($7+0,5+0,5$) Tagen des Berichtsmonats keine Produktion statt.

2. Betriebszeiten

Zu dem Personal zählen alle Arbeitnehmer, die entsprechend dem Schwerpunkt ihrer Tätigkeit in dem jeweiligen Produktionsbereich beschäftigt sind.

Ein Arbeitsplatz muß für mindestens 50 % der Werktage des Berichtsmonats eingerichtet worden sein.

Die folgenden Beispiele stellen mögliche Verfahren zur Berechnung der Betriebszeit dar.

Einschichtbetrieb:

Der Betrieb XYZ produziert in jeder Woche über 50 Stunden. Bei 4 gleichen Wochen beträgt die monatliche Betriebszeit 200 Stunden.

Mehrschichtbetrieb:

Der Betrieb XYZ produziert täglich in 3 Schichten zu je 7 Stunden. Bei 20 Werktagen eines Berichtsmonats beträgt die monatliche Betriebszeit 420 Stunden.

Zweck, Art und Umfang der Erhebung

Die Erhebung erfolgt im Auftrag des Bundesministers für Wirtschaft. Sie erstreckt sich auf höchstens 10.000 Betriebe im Verarbeitenden Gewerbe, die von der amtlichen Produktionsstatistik erfaßt werden. Sie soll Informationen über die Anzahl der monatlichen Produktionslage und Betriebszeiten bereitstellen. Dies dient einer besseren Erfassung konjunktureller Einflüsse sowie einer besseren Berechnung von Produktivitätskennzahlen.

Rechtsgrundlage

§ 7 Abs. 1 des Gesetzes über die Statistik für Bundeszwecke (Bundesstatistikgesetz - BStatG) vom 22. Januar 1987 (BGBl. I S. 462, 565), geändert durch Artikel 2 des Gesetzes vom 17. Dezember 1990 (BGBl. I S. 2837). Die Auskunftserteilung ist freiwillig.

Geheimhaltung

Die erhobenen Einzelangaben werden nach § 16 BStatG ausnahmslos geheimgehalten. Sie dienen ausschließlich statistischen Zwecken.

Hilfsmerkmale, Löschen und Ordnungsnummern

Name und Telefonnummer der für Rückfragen zur Verfügung stehenden Personen sind Hilfsmerkmale, die lediglich der technischen Durchführung der Erhebung dienen. Sie werden zusammen mit dem Erhebungsvordruck nach Abschluß der Prüfung der Angaben auf Vollständigkeit und Schlüssigkeit vernichtet. Die Betriebsnummer dient der rationellen Aufbereitung der Ergebnisse und der Identifikation des Betriebes. Mit Hilfe der SYPRO-Nummer kann der Betrieb einem Wirtschaftszweig entsprechend dem Schwerpunkt seiner Tätigkeit zugeordnet werden.

Erhebungsbogen Version B – Vorderseite

Statistisches Landesamt	Erfassung von Produktionstagen und Betriebszeiten im Verarbeitenden Gewerbe	
	Betriebs-Nr.: <input type="text"/> Bei Rückfragen bitte angeben.	
Unsere Telefon-Nr.:	Bitte senden Sie 1 Exemplar bis zum 5. des folgenden Monats zurück.	
Name und Anschrift des Betriebes:	Rechtsgrundlagen, Erläuterungen und Beispiele finden Sie im Anschluß an die Fragen.	
	Name eines Ansprechpartners:	
	<input type="text"/>	
	Telefon-Nr.:	
	<input type="text"/>	
Betriebs-Nr. (SST 1 - 9):	SYPRO-Nr.:	Berichtsmonat:
<input type="text"/>	<input type="text"/>	<input type="text"/>

Beachten Sie bitte die Erläuterungen und Beispiele auf der Rückseite!

1. Berechnung der Produktionstage

Für die Berechnung dieses Merkmals werden die Tage eines Berichtsmonats, an denen keine oder nur eine reduzierte Produktion stattfand, benötigt.

Es fand an Tagen des Berichtsmonats keine Produktion statt.

2. Berechnung der Betriebszeiten

Mit der Betriebszeit ist die monatliche Stundenzahl gemeint, in der der Produktionsbetrieb in Gang gewesen ist. Zur Betriebszeit zählen auch Vorbereitungs- bzw. Hilfszeiten.

Zur Berechnung der Betriebszeit werden folgende Angaben benötigt:

SYPRO-Nr. des Betriebsteils	Schichtdauer in Stunden		Schichtpersonal		Anzahl der Tage
	1. Schicht	übrige Schichten	1. Schicht	übrige Schichten	
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Erhebungsbogen Version B – Rückseite

Erläuterungen und Beispiele

1. Produktionstage

Der Betrieb XYZ produziert in seiner Produktionsabteilung täglich in 8 Stunden mit 200 Arbeitnehmern.

An einem Monat mit 31 Kalendertagen wurde an 3 Samstagen und 4 Sonntagen nicht produziert. Es entstanden damit 7 fehlende Tage.

An einem Freitag arbeiteten nur 100 Arbeitnehmer. Der Freitag ist daher als ein halber Fehtag zu werten.

An einem Dienstag wurde nur in 4 Stunden produziert. Der Ausfall von 4 Stunden ist daher als ein halber Fehtag zu werten.

Im Betrieb XYZ fand damit an 8 ($7+0,5+0,5$) Tagen des Berichtsmonats keine Produktion statt.

2. Betriebszeiten

Zu dem Personal zählen alle Arbeitnehmer, die entsprechend dem Schwerpunkt ihrer Tätigkeit in dem jeweiligen Produktionsbereich beschäftigt sind.

Falls die Anzahl der Arbeitnehmer im Verlauf eines Monats wechselt, können in der Spalte "Personal" Durchschnittswerte eingesetzt werden.

Falls im Berichtsmonat die Schichtdauern täglich variieren, genügt die Angabe von Durchschnittswerten.

Die folgende Tabelle zeigt die Daten des Betriebes XYZ für seine Betriebsteile im Einschicht- und im Mehrschichtbetrieb:

SYPRO-Nr. des Betriebsteils	Schichtdauer in Stunden		Schichtpersonal		Anzahl der Tage
	1. Schicht	übrige Schichten	1. Schicht	übrige Schichten	
3240	8	15	35	55	22
3280	8		12		20

Zweck, Art und Umfang der Erhebung

Die Erhebung erfolgt im Auftrag des Bundesministers für Wirtschaft. Sie erstreckt sich auf höchstens 10.000 Betriebe im Verarbeitenden Gewerbe, die von der amtlichen Produktionsstatistik erfaßt werden. Sie soll Informationen über die Anzahl der monatlichen Produktionstage und Betriebszeiten bereitstellen. Dies dient einer besseren Erfassung konjunktureller Einflüsse sowie einer besseren Berechnung von Produktivitätskennzahlen.

Rechtsgrundlage

§ 7 Abs. 1 des Gesetzes über die Statistik für Bundeszwecke (Bundesstatistikgesetz - BStatG) vom 22. Januar 1987 (BGBl. I S. 462, 565), geändert durch Artikel 2 des Gesetzes vom 17. Dezember 1990 (BGBl. I S. 2837). Die Auskunftserteilung ist freiwillig.

Geheimhaltung

Die erhobenen Einzelangaben werden nach § 16 BStatG ausnahmslos geheimgehalten. Sie dienen ausschließlich statistischen Zwecken.

Hilfsmerkmale, Löschen und Ordnungsnummern

Name und Telefonnummer der für Rückfragen zur Verfügung stehenden Personen sind Hilfsmerkmale, die lediglich der technischen Durchführung der Erhebung dienen. Sie werden zusammen mit dem Erhebungsvordruck nach Abschluß der Prüfung der Angaben auf Vollständigkeit und Schlüssigkeit vernichtet. Die Betriebsnummer dient der rationellen Aufbereitung der Ergebnisse und der Identifikation des Betriebes. Mit Hilfe der SYPRO-Nummer kann der Betrieb einem Wirtschaftszweig entsprechend dem Schwerpunkt seiner Tätigkeit zugeordnet werden.

Erhebungsbogen Version C – Vorderseite

Statistisches Landesamt		Erfassung von Produktionstagen und Betriebszeiten im Verarbeitenden Gewerbe	
Unsere Telefon-Nr.:		Betriebs-Nr.: Bei Rückfragen bitte angeben.	<input type="text"/>
Name und Anschrift des Betriebes:		Bitte senden Sie 1 Exemplar bis zum 5. des folgenden Monats zurück.	
		Rechtsgrundlagen, Erläuterungen und Beispiele finden Sie im Anschluß an die Fragen.	
		Name eines Ansprechpartners:	
		<input type="text"/>	
		Telefon-Nr.:	
		<input type="text"/>	
Betriebs-Nr. (SST 1 - 9):	SYPRO-Nr.:	Berichtsmonat:	
<input type="text"/>	<input type="text"/>	<input type="text"/>	

Beachten Sie bitte die Erläuterungen und das Beispiel auf der Rückseite!

[illegible]

Erhebungsbogen Version C – Rückseite

Erläuterungen

1. Für die Berechnung der Produktionstage und Betriebszeiten ist eine korrekte Numerierung der Betriebsteile erforderlich.
2. Wenn in einem Betriebsteil das Schichtpersonal oder die Schichtdauer zwei- bis dreimal im Monat variieren, müssen mehrere Angaben unter der gleichen SYPRO-Nummer (Spalte 1) gemacht werden (Siehe hierzu auch das Beispiel.)
3. Wenn die Anzahl des Schichtpersonals oder die Schichtdauer in einem Betriebsteil täglich variieren, genügt die Angabe von Durchschnittswerten.
4. Das Personal ist entsprechend dem Schwerpunkt seiner Tätigkeit dem Betriebsteil zuzuordnen. Es sind alle Arbeitnehmer, die für die Produktion benötigt werden, aufzuführen.
5. Ganztägige Produktionsunterbrechungen dürfen nicht unter den Tagen (Spalte 4) aufgeführt werden.

Ausfüllbeispiel

Betriebsteil	Schichtdauer				Schichtpersonal				Tage
1	2				3				4
SYPRO-Nr.	Schicht 1	Schicht 2	Schicht 3	Schicht 4	Schicht 1	Schicht 2	Schicht 3	Schicht 4	Anzahl
3240	8	9	6		35	35	20		15
3240	8	9	4		35	35	10		5
3257	8	8			40	40			22
3280	8				12				20

Zweck, Art und Umfang der Erhebung

Die Erhebung erfolgt im Auftrag des Bundesministers für Wirtschaft. Sie erstreckt sich auf höchstens 10.000 Betriebe im Verarbeitenden Gewerbe, die von der amtlichen Produktionsstatistik erfaßt werden. Sie soll Informationen über die Anzahl der monatlichen Produktionstage und Betriebszeiten bereitstellen. Dies dient einer besseren Erfassung konjunktureller Einflüsse sowie einer besseren Berechnung von Produktivitätskennzahlen.

Rechtsgrundlage

§ 7 Abs. 1 des Gesetzes über die Statistik für Bundeszwecke (Bundesstatistikgesetz - BStatG) vom 22. Januar 1987 (BGBl. I S. 462, 565), geändert durch Artikel 2 des Gesetzes vom 17. Dezember 1990 (BGBl. I S. 2837). Die Auskunftserteilung ist freiwillig.

Geheimhaltung

Die erhobenen Einzelangaben werden nach § 16 BStatG ausnahmslos geheimgehalten. Sie dienen ausschließlich statistischen Zwecken.

Hilfsmerkmale, Löschen und Ordnungsnummern

Name und Telefonnummer der für Rückfragen zur Verfügung stehenden Personen sind Hilfsmerkmale, die lediglich der technischen Durchführung der Erhebung dienen. Sie werden zusammen mit dem Erhebungsvordruck nach Abschluß der Prüfung der Angaben auf Vollständigkeit und Schlüssigkeit vernichtet. Die Betriebsnummer dient der rationalen Aufbereitung der Ergebnisse und der Identifikation des Betriebes. Mit Hilfe der SYPRO-Nummer kann der Betrieb einem Wirtschaftszweig entsprechend dem Schwerpunkt seiner Tätigkeit zugeordnet werden.

4 Durchführung der Befragung

Die Interviewer wurden in drei ca. zweistündigen Schulungssitzungen auf ihre Aufgabe bei diesem Projekt vorbereitet. Besonderes Augenmerk wurde bei der Schulung auf die Ermittlung der Zielperson in dem Betrieb gerichtet, da dies eine der schwierigsten Phasen der Befragung sein würde. Durch Vermittlung von Argumentationshilfen und Rollenspiele wurde den Interviewern verdeutlicht, wie es ihnen gelingen würde, hartnäckige Chefsekretärinnen und ähnliche „Hindernisse“ auf dem Weg zur Zielperson zu umgehen und mit welchen Argumenten sie sich letztendlich auch der Zielperson würden „verkaufen“ können.

Die Befragung wurde im Herbst 1992 durchgeführt. Die Aufgabe der Interviewer bestand zunächst darin, bei vorgegebenen Firmenadressen anzurufen und zu ermitteln, wer dort zu Fragen von Produktionstagen und Betriebszeiten kompetent Auskunft geben könnte. Mit dieser Person sollte dann Kontakt aufgenommen und ein Termin vereinbart werden, zu dem der Interviewer den „Selbstaussfüllerbogen“ vorbeibringen, die Zielperson beim Ausfüllen des Bogens „beobachten“ und anschließend befragen sollte. Falls dieses optimale Vorgehen nicht zu realisieren sein sollte, hatte der Interviewer den Erhebungsbogen bei der Zielperson zu belassen und einen zweiten Termin zu vereinbaren, zu dem der Bogen abgeholt und die Nachbefragung realisiert werden könnte.

Insgesamt sind 407 Firmenadressen ausgegeben worden; davon konnten 20 Firmen nicht erreicht werden, in 14 Fällen existierte die Firma nicht (mehr), in 29 Fällen hatten die Firmen weniger als 100 (respektive 75) Beschäftigte; daneben gab es 30 weitere Firmen, die als sogenannte „neutrale“ Ausfälle (Firma produziert nicht/nicht mehr) aus allen folgenden Überlegungen ausgeschlossen werden konnten.

Setzt man die verbleibenden 314 Firmen auf 100 %, so ergibt sich folgende Ausschöpfung der Stichprobe (siehe folgende Übersicht 1), wobei, wie zuvor bemerkt, die Frage der Ausschöpfung hier allenfalls von deskriptivem Interesse ist.

Übersicht 1
Ausschöpfung der Stichprobe

Bereinigte Stichprobe (Netto-Adressen) (insgesamt 314 Firmen = 100 %)	Absolut	%
Nichtneutrale Ausfälle		
– Kontaktperson verweigert Zugang zur Zielperson	12	3,8
– Zielperson nicht erreicht	43	13,7
– Zielperson legt auf	1	0,3
– Zielperson hat keine Zeit	67	21,3
– Zielperson hat keine Lust, kein Interesse	59	18,8
– Zielperson verweigert explizit	47	15,0
Nichtneutrale Ausfälle insgesamt	229	72,9
Prinzipiell teilnahmebereit	85	27,1

Die prinzipiell Teilnahmebereiten setzen sich zusammen aus

- (1) 69 Personen, die bereit waren, einen Interviewer im Betrieb zu empfangen und
- (2) 16 Personen, die dazu nicht bereit waren, aber angaben, sie würden den Erhebungsbogen, so er denn zugeschickt oder zugefaxt würde, ausfüllen und sich telefonisch nachbefragen lassen.

Von den 16 Personen, die sich den Bogen hatten zuschicken lassen, sandten ihn 11 auch ausgefüllt zurück und ließen sich telefonisch nachbefragen. Die restlichen 5 erklärten sich nicht bereit, den Bogen auszufüllen, nachdem er ihnen vorgelegen hatte, bzw. erklärten sich nicht bereit, an der Nachbefragung teilzunehmen. Zum Kontakt vor Ort kam es tatsächlich in 69 Fällen, wobei in 44 Fällen ein Besuch und in 25 Fällen zwei Besuche nötig waren.

Da der Erhebungsbogen sehr kurz war und nicht so viel Aufwand beim Ausfüllen erforderte, wie ursprünglich angenommen, konnte er häufiger als erwartet gleich ausgefüllt werden. Von den insgesamt 69 vollständig und im Sinne des Konzepts der Studie vor Ort realisierten Befragungen wurden 20 mit Version A des Erhebungsbogens durchgeführt, 23 mit Version B und 26 mit Version C.

Bei der Bewertung der quantitativen Ergebnisse ist zu berücksichtigen, daß aufgrund der relativ geringen Fallzahlen statistische Standardverfahren nicht greifen. Eine quantitative Analyse der Ergebnisse war aber auch nicht Gegenstand und Ziel des Pretests. Ziel des Pretests war vielmehr die qualitative Analyse; Tendenzen, die sich aus der quantitativen Analyse ergeben, können die Ergebnisse der qualitativen Analyse stützen.

Das Übergewicht der Firmen im SYPRO-Zweisteller 32 in der Stichprobe spiegelt sich auch in der Verteilung der realisierten Fälle wider. Während die anderen Branchen mit jeweils 10 bis 12 Fällen in die Ergebnisse eingehen, gehören 26 der realisierten Firmen dem SYPRO-Zweisteller 32 an. Wie aus der folgenden Tabelle 1 ebenfalls zu ersehen ist, ist es gelungen, die verschiedenen Versionen des Erhebungsbogens in allen Branchen zum Einsatz zu bringen:

‘Tabelle 1
Einsatz der drei Versionen des Erhebungsbogens nach Branchen

Branche (in Klammern: SYPRO-Zweisteller)	Anzahl der eingesetzten Fragebogen			
	Fragebogen insgesamt	Fragebogen-Versionen		
		A	B	C
Gewinnung und Verarbeitung von Steinen und Erden (25)	10	2	5	3
Maschinenbau (32)	26	7	7	12
Elektrotechnik, Reparatur von elektrischen Geräten für den Haushalt (36)	10	2	4	4
Chemische Industrie (40)	11	3	4	4
Ernährungsgewerbe (68)	12	6	3	3
Insgesamt ...	69	20	23	26

54 (78,3 %) der 69 Auskunftspersonen in den Betrieben waren Männer, 14 (20,3 %) waren Frauen (für eine Person war das Geschlecht nicht angegeben); nach Schätzung der Interviewer waren 19 Befragte (27,5 %) unter 35 Jahre alt, 30 (43,5 %) zwischen 35 und 50 Jahre und 15 (21,7 %) älter als 50 Jahre (für 5 Personen lag keine Altersschätzung vor). Mehr als die Hälfte der Auskunftspersonen hatte einen Hochschulabschluß ohne (47,8 % = 33) oder mit Promotion (4,3 % = 3); 30,4 % (21 Personen) der Befragten hatte nach eigenen Angaben eine Lehre abgeschlossen, 15,9 % (11) hatten einen Fachhochschulabschluß, 7,2 % (5) die Meisterprüfung (hier wurden Mehrfachangaben gemacht). 21,7 % (15 Personen) bezeichneten ihre Position im Betrieb als „Führungskraft, an die Führungskräfte berichten“; 63,7 % (44) sind nach eigener Auskunft „Führungskraft mit Personalverantwortung“, 6 (8,7 %) stufen sich als „Technische Angestellte“ ein, vier Personen (5,7 %) wurden unter „Anderes“ eingeordnet.

5 Ergebnisse

Die Ergebnisse des Pretests werden in drei Teilen dargestellt. In einem ersten Teil wird beschrieben, wie die Auskunftspersonen den Erhebungsbogen ausgefüllt haben. Der zweite Teil, der Hauptteil der qualitativen Analyse, dokumentiert die Bewertung des Verfahrens und des Erhebungsbogens aus der Sicht der Auskunftspersonen als Ergebnis der Nachbefragung. Im dritten Teil schließlich sollen Auszüge aus den Berichten der Interviewer einen kurzen Einblick in die „Atmosphäre“ der Befragungssituation verschaffen.

5.1 Die Ausfüllqualität der Erhebungsbogen

Zuerst soll aufgezeigt werden, in welchen Umfang die Erhebungsbogen ausgefüllt wurden.

Version A:

17 der 20 Bogen aus Version A (85 %) enthalten Angaben über die Produktionstage, 19 (95 %) Angaben über die monatliche Betriebszeit, 18 (90 %) zum Personal und 16 (80 %) zu den Arbeitsplätzen. Der Bogen ist weitgehend vollständig ausgefüllt worden.

Version B:

21 der 23 Bogen (91,3 %) enthalten Angaben zu den Produktionstagen, ebenso zur „Schichtdauer 1. Schicht“; 17 Bogen (73,9 %) enthalten Angaben zu „Schichtpersonal 1. Schicht“, 19 (82,6 %) zur „Anzahl der Tage“. Die übrigen Angaben der Tabelle („Übrige Schichten“) liegen knapp unter 50 %, was aber daran liegen könnte, daß es faktisch keine „übrige Schicht“ gibt.

Version C:

20 von 26 Bogen enthalten Angaben zur Anzahl der Tage. Die Werte der Angaben zur Schichtdauer fallen von Schicht 1 (24 = 92,3 %) über Schicht 2 (15 = 57,7 %) und Schicht 3 (7 = 26,9 %) zu Schicht 4 (2 = 7,7 %), was ebenfalls durch faktisch nicht vorhandene Schichten zu erklären ist. Entsprechend fehlen dann auch die Angaben zum Schichtpersonal.

Alle Bogen wurden weitgehend vollständig ausgefüllt; aus der formalen Vollständigkeit der ausgefüllten Bogen ist also, betrachtet man das Gesagte, keine Entscheidung zugunsten einer der drei zu testenden Erhebungsversionen zu treffen. Es wird also darauf ankommen, wie die Auskunftspersonen sich zu den jeweils getesteten Versionen äußern.

5.2 Das Urteil der Auskunftspersonen

Das Urteil der Auskunftspersonen wurde mit Hilfe einer kurzen Befragung ermittelt, welche die Interviewer vor Ort nach Abschluß des Ausfüllens des Erhebungsbogens mit der Auskunftsperson durchführten. Der dazu eingesetzte Fragebogen lag in drei Versionen vor – analog der drei Versionen des Erhebungsbogens – wobei die Fragebogen für die Versionen A und B identisch waren.

5.2.1 Verständlichkeit der Berechnung der Produktionstage (nur Versionen A und B)

Die Berechnung der Produktionstage erfolgt nur in den Versionen A und B mit Hilfe einer eigenen Frage; Version C enthält diese Frage nicht.

Von den 43 Auskunftspersonen der Versionen A und B bezeichneten 16 (37,2 %) den Fragetext zu den Produktionstagen als „sehr verständlich“, 15 (34,9 %) als „einigermaßen verständlich“, 10 (23,3 %) als „wenig verständlich“ und nur 2 (4,7 %) als „überhaupt nicht verständlich“. Obwohl der Fragetext in den beiden Versionen identisch war, schneidet Version A mit 85 % „verständlich“-Nennungen deutlich besser ab als Version B mit 60,8 %. Dies liegt wohl weniger an der Version des Bogens an sich als an der Tatsache, daß – wie später gezeigt wird – die Erläuterungen zu dieser Frage auf der Rückseite der Bogen bei Version A positiver bewertet werden als bei Version B. Personen, die angeben, der Text sei „wenig verständlich“ oder „überhaupt nicht verständlich“, wurden gefragt, was unverständlich gewesen sei bzw. was verbessert werden könnte. Dabei kam es zu folgenden Kritikpunkten bzw. Verbesserungsvorschlägen:

- Mehrschicht-Betriebe mit unterschiedlichen Schichtzeiten können mit dem Bogen nicht erfaßt werden (Version A: eine Nennung).
- Schwierigkeiten, wenn z. T. keine vollen Schichten gearbeitet werden (Version B: eine Nennung).
- Unklar ist, wie mit Sonntagen oder Feiertagen umzugehen sei (Version A: eine Nennung, Version B: eine Nennung).
- Die negative Frageformulierung entspricht nicht der betrieblichen Realität (Version A: eine Nennung, Version B: 6 Nennungen): „Die Frage sollte eher umgekehrt gestellt werden, d. h., wie viele Produktionstage angefallen sind“. Oder: „Eigentlich wird nach den Produktionstagen gefragt; nicht sinnvoll, nach Negation zu fragen, entgegen der Praxis, unlogisch“.
- Die Sprache ist zu kompliziert (Version B: eine Nennung).
- Der Begriff „Produktionstage“ ist nicht eindeutig (Version B: zwei Nennungen).

Die Hauptkritik an dem Fragetext zu „Produktionstagen“ bezieht sich also auf die negative Formulierung der Frage, die offensichtlich der innerbetrieblichen Denkweise nicht entspricht: Statt nach Tagen zu fragen, an denen keine Produktion stattfand, sollte man besser

nach Tagen fragen, an denen produziert wurde. Zu überlegen wäre, ob der Begriff „Produktionstage“ nicht durch einen anderen (vorgeschlagen wurde u. a. „Arbeitstage“ oder „Werkstage“) ersetzt werden könnte.

5.2.2 Aufwand für die Befragten beim Beantworten der Frage nach den Produktionstagen (nur Versionen A und B)

Trotz der Kritik an der Frage nach den Produktionstagen scheint ihre Beantwortung nicht sonderlich schwierig oder aufwendig gewesen zu sein. Von den 43 Auskunftspersonen der Versionen A und B haben 31 (72,1 %) diese Frage allein beantwortet, 12 (27,9 %) haben andere Personen hinzugezogen, wobei in 9 Fällen eine weitere Person konsultiert wurde; in einem Falle wurden zwei und in zwei Fällen drei weitere Personen hinzugezogen.

Um die benötigten Unterlagen oder Informationen zum Beantworten der Frage zu beschaffen, mußte allerdings häufiger auf andere Personen im Betrieb zurückgegriffen werden. Bei Version A war dies 13mal (65,0 %) der Fall, bei Version B nur 9mal (39,1 %).

In 8 von 43 Fällen (18,6 %) mußten zur Beantwortung der Frage Informationen oder Unterlagen erst neu zusammengestellt werden, in den anderen 35 Fällen (81,4 %) lag alles Benötigte im Betrieb bereits vor. Dementsprechend kurz war die Zeit, die von den meisten Auskunftspersonen zur Beantwortung der Frage gebraucht wurde (siehe hierzu die folgende Tabelle 2).

Tabelle 2
Zeitlicher Aufwand zur Beantwortung der Frage
nach Produktionstagen

Dauer der Informationsbeschaffung (Berechnung Produktionstage)	Fragebogen-Versionen		
	A+B	A	B
	(N = 43)	(N = 20)	(N = 23)
Anzahl der Nennungen (in Klammern in %)			
bis 5 Minuten	24 (55,8)	10 (50,0)	14 (60,9)
6 bis 10 Minuten	5 (11,6)	2 (10,0)	3 (13,0)
11 bis 20 Minuten	4 (9,3)	1 (5,0)	3 (13,0)
21 bis 59 Minuten	3 (7,0)	2 (10,0)	1 (4,3)
1 Stunde und mehr	1 (2,3)	1 (5,0)	0 (0,0)
Keine Angabe	6 (13,9)	4 (20,0)	2 (8,7)

Die Hälfte (Version A) bzw. mehr als die Hälfte (Version B) aller Auskunftspersonen benötigt höchstens 5 Minuten zum Beantworten der Frage. Nur eine einzige Person gab an, mehr als eine Stunde dafür gebraucht zu haben.

5.2.3 Probleme beim Ausfüllen der Berechnung der Produktionstage (nur Versionen A und B)

Am Ende des Blocks „Berechnung der Produktionstage“ wurde den Auskunftspersonen eine Liste mit Sachverhalten vorgelegt, und sie wurden gefragt, ob dieser Sachverhalt in ihrer Firma zuträfe und – falls ja – ob sich daraus Probleme beim Beantworten der Frage ergeben hätten.

Beim Ausfüllen der „Berechnung der Produktionstage“ war es

- für 5 Personen (11,6 %) ein Problem, daß es mehr als vier Wechsel bei den Schichtsystemen im Betrieb gibt,
- für 11 Personen (25,6 %) ein Problem, daß es in dem Betrieb ständig wechselnde Personalstärken in den einzelnen Schichten gibt,
- für 19 Personen (44,2 %) ein Problem, daß kein EDV-Programm zur Berechnung des Merkmals vorhanden war,
- für 10 Personen (23,3 %) ein Problem, daß zu viele Produktionsabteilungen im Unternehmen sind,
- für 4 Personen (9,3 %) ein Problem, daß keine Unterscheidung zwischen Produktions- und Verwaltungsabteilungen besteht und
- für 3 Personen (7,0 %) ein Problem, daß die Art der Produktion in ihrem Betrieb eine Erfassung des Merkmals nicht zuläßt.

Das Bearbeiten der Frage nach den Produktionstagen ist also im Prinzip zwar einfach und wenig aufwendig, kann aber unter bestimmten betrieblichen Konstellationen (vor allem ohne entsprechendes EDV-Programm, ständig wechselnde Schichten, zu viele Produktionsabteilungen im Betrieb) doch problembelastet sein. Es stellt sich hier die Frage, ob das Instrument zur Erfassung der Produktionstage komplex genug ist, betriebliche Realität zu erfragen bzw. wie das Instrument gestaltet werden muß, damit es die betriebliche Realität besser abbildet.

5.2.4 Verständlichkeit der Berechnung der Betriebszeiten (Versionen A und B sowie Gesamtbogen Version C)

Im Gegensatz zu den Versionen A und B besteht Version C aus einer einzigen Tabelle. Sie wird im folgenden der Tabelle 2 aus Version A und Version B gegenübergestellt.

Als „sehr verständlich“ bezeichneten 45 % (9 Personen) Tabelle 2 in Version A, in Version B waren es 30,4 % (7 Personen); Version C bezeichneten 11,5 % (3 Personen) als „sehr verständlich“. Nimmt man die Kategorien „sehr verständlich“ und „einigermaßen verständlich“ zusammen, so ergeben sich für A 65,0 %, für B 56,5 %, für C sogar 69,2 %.

Die Verständlichkeit ist also bei A und C besser als bei B, wobei A gegenüber C einen deutlichen Vorteil in der Extremkategorie „sehr verständlich“ hat.

Bei Version A wurde kritisiert (hier sollten die Zielpersonen angeben, was am Text unverständlich ist), daß

- der Begriff „Betriebszeit“ nicht genau definiert sei,
- der Begriff „Betriebszeit“ überhaupt verwendet wurde, während in anderen Statistiken von Arbeitszeit gesprochen würde,
- der Begriff „Arbeitsplatz“ nicht gut definiert sei: „Handelt es sich um vorhandene oder besetzte Plätze?“.

Sollte Version A beibehalten werden, müßten also zumindest die beiden Begriffe „Betriebszeiten“ und „Arbeitsplatz“ sorgfältig definiert werden.

Bei Version B wurde kritisiert, daß

- die monatlichen Stundenzahlen nicht erfaßt würden,
- der Begriff „Betriebszeit“ nur eine komplizierte Umschreibung des Begriffes „monatliche Arbeitszeit“ sei; man solle besser nach den gesamt geleisteten Arbeitsstunden fragen, weil dadurch auch Teilzeitkräfte erfaßt würden,
- die Begriffe „1. Schicht“ und „übrige Schichten“ unklar seien.

Auch hier also wird in erster Linie die unklare Formulierung zentraler Begriffe der Fragestellung kritisiert; sie zu präzisieren wäre erforderlich, wenn man sich für Version B entscheiden würde.

Bei Version C wurde kritisiert, daß

- das Schema besser erklärt werden müßte,
- bestimmte Aspekte nicht ausreichend definiert seien, z. B. „Betriebsteile“,
- nicht klar sei, worauf sich Spalte 4 „Tage“ bezieht.

Wollte man Version C weiterverwenden, müßte man also deutlich machen, worauf sich die Spalte „Tage“ bezieht (auch wenn darauf über den „Berichtsmonat“ geschlossen werden könnte), den Begriff „Betriebsteile“ besser definieren und das Schema insgesamt besser erklären.

5.2.5 Aufwand beim Ausfüllen der Tabelle Betriebszeiten

Während im Bogen bei Version C und bei Version A die Frage der Berechnung der Betriebszeiten von den Auskunftspersonen weitgehend alleine, d. h. ohne Hinzuziehen dritter Personen beantwortet werden konnte (C: 80,8 %, A: 75,0 %), war dies bei Version B bei nur knapp der Hälfte aller Fälle möglich.

Um die benötigten Unterlagen oder Informationen zum Beantworten der Frage zu beschaffen, mußte – wie bei den Produktionstagen auch – häufiger auf andere Personen zurückgegriffen werden. Bei Version A war dies 13mal (65 %) der Fall, bei Version B 10mal (43,5 %) und bei Version C 16mal (61,5 %). Wie bei den Produktionstagen muß also auch bei den Betriebszeiten bei Version B am wenigsten an Information bzw. Unterlagen

beschafft werden. Während bei den Versionen A und C für jeweils ca. 75 % aller Fälle die benötigten Informationen im Betrieb bereits vorlagen, war dies bei Version B nur in 56,5 % der Fälle gegeben.

Bei knapp der Hälfte aller Fälle bei Versionen A und B benötigten die Befragten zur Berechnung der Betriebszeiten weniger als 5 Minuten.

Tabelle 3
Zeitlicher Aufwand/Dauer zur Berechnung der Betriebszeiten
bei den Erhebungsbogen-Versionen A und B

Dauer der Informationsbeschaffung (Berechnung Produktionstage)	Fragebogen-Versionen		
	A+B (N = 43)	A (N = 20)	B (N = 23)
	Anzahl der Nennungen (in Klammern in %)		
bis 5 Minuten	20 (46,5)	9 (45,0)	14 (47,8)
6 bis 10 Minuten	4 (9,3)	1 (5,0)	3 (13,0)
11 bis 20 Minuten	5 (11,6)	2 (10,0)	3 (13,0)
21 bis 59 Minuten	2 (4,7)	1 (5,0)	1 (4,3)
1 Stunde und mehr	1 (2,3)	1 (5,0)	0 (0,0)
Keine Angabe	11 (25,6)	6 (30,0)	5 (21,7)

Die Angaben in dieser Tabelle belegen insgesamt sehr anschaulich, daß das Ausfüllen der Angaben zu den Betriebszeiten keine großen zeitlichen Probleme mit sich bringt. Schwierigkeiten gibt es eher wieder bei betrieblichen Strukturmerkmalen.

5.2.6 Probleme beim Ausfüllen: „Betriebszeiten“

Die meisten Probleme beim Ausfüllen der „Betriebszeiten“ gibt es, weil es in Unternehmen zu viele Produktionsabteilungen gibt (20 % oder 4 Nennungen bei Version A; 26,1 % oder 6 Nennungen bei Version B) und weil die Anzahl des Personals in den einzelnen Schichten ständig wechselt (15 % oder 3 Nennungen bei Version A; 26,1 % oder 6 Nennungen bei Version B).

Eine Person in Version A und drei Personen in Version B hatten Probleme mit mehr als viermaligem Wechseln im Schichtsystem. Daß kein EDV-Programm zur Steuerung des Merkmals vorhanden war, war für 2 (8,7 %) Personen in Version B ein Problem; in Version A trat dieses Problem nicht auf.

In Version C machte kein einziger der beschriebenen Sachverhalte Probleme beim Ausfüllen des Bogens. Dies deutet darauf hin, daß Version C diejenige ist, die der Komplexität der betrieblichen Realität am nächsten kommt oder zumindest die Erfassung der Abläufe am ehesten ermöglicht.

5.2.7 Gesamtdauer der Bearbeitung

Wenngleich Version C diejenige mit der detaillierteren Abfragestruktur ist, benötigen doch 10 von 26 Befragten (38,5 %) höchstens 5 Minuten, um den Bogen auszufüllen. Bei den Versionen A und B benötigen die Auskunftspersonen relativ am häufigsten eine Zeitspanne von 6 bis 10 Minuten.

Auf der anderen Seite gilt für alle Versionen, daß etwa die Hälfte aller Bogen in höchstens 10 Minuten ausgefüllt werden kann. Allenfalls der Spitzenwert bei Version C könnte dieser Version unter Zeitaspekten einen Vorteil verschaffen; keine Unterschiede gibt es dagegen bei den Fällen, die länger als 10 Minuten zum Ausfüllen beanspruchen.

Die Angaben zur Schichtdauer sind nach Darlegung der Auskunftspersonen in der C-Version zu ca. drei Viertel genaue Werte; bei den Angaben zum Schichtpersonal halten sich genaue Werte und Durchschnittswerte in etwa die Waage.

Tabelle 4
Gesamtdauer der Bearbeitungszeit des Erhebungsbogens

Dauer der Informationsbeschaffung (Produktionstage und Betriebszeiten bei A und B; Gesamtzeit bei C)	Fragebogen-Versionen		
	A	B	C
	(N = 20)	(N = 23)	(N = 26)
	Anzahl der Nennungen (in Klammern in %)		
bis 5 Minuten	2 (10,0)	3 (13,0)	10 (38,5)
6 bis 10 Minuten	6 (30,0)	9 (39,1)	2 (7,7)
11 bis 20 Minuten	2 (10,0)	4 (17,4)	4 (15,4)
21 bis 59 Minuten	4 (20,0)	4 (17,4)	4 (15,4)
1 Stunde und mehr	2 (10,0)	1 (4,3)	2 (7,7)
Keine Angabe	4 (20,0)	2 (8,7)	4 (15,4)

5.3 Bewertung der Bogen insgesamt

Bei der Globalbewertung des Erhebungsbogens nach Einfachheit, Interessantheit und Übersichtlichkeit zeigt sich (siehe Tabelle 5, S. 109), daß

- Version A als einfach auszufüllen und übersichtlich bewertet wird, aber als uninteressant,
- Version B zwar etwas schwieriger auszufüllen und etwas weniger übersichtlich ist als A und C, dafür noch am ehesten interessant ist,
- Version C einfach auszufüllen ist, einigermaßen übersichtlich, aber ebenfalls nur wenig interessant.

Tabelle 5
Gesamtbewertung des Erhebungsbogens

Allgemeine Bewertung	Fragebogen-Versionen		
	A (N = 20)	B (N = 23)	C (N = 26)
	Anzahl der Nennungen (in Klammern in %)		
<i>Das Ausfüllen des Bogens war:</i>			
sehr einfach	4 (20,0)	3 (13,0)	7 (26,9)
eher einfach	11 (55,0)	11 (47,8)	12 (46,2)
eher nicht einfach	4 (20,0)	4 (17,4)	4 (15,4)
überhaupt nicht einfach	1 (5,0)	4 (17,4)	1 (3,8)
Keine Angabe	0 (0,0)	1 (4,3)	2 (7,7)
<i>Das Ausfüllen des Bogens war:</i>			
sehr interessant	0 (0,0)	1 (4,3)	0 (0,0)
eher interessant	3 (15,0)	5 (21,7)	2 (7,7)
eher nicht interessant	8 (40,0)	5 (21,7)	12 (46,2)
überhaupt nicht interessant	9 (45,0)	10 (43,5)	10 (38,5)
Keine Angabe	0 (0,0)	2 (8,7)	2 (8,7)
<i>Von der graphischen Gestaltung her war der Bogen:</i>			
sehr übersichtlich	6 (30,0)	5 (21,7)	7 (26,9)
einigermaßen übersichtlich	13 (65,0)	12 (52,2)	13 (50,0)
eher nicht übersichtlich	1 (5,0)	5 (21,7)	4 (15,4)
überhaupt nicht übersichtlich	0 (0,0)	1 (4,3)	0 (0,0)
Keine Angabe	0 (0,0)	0 (0,0)	2 (7,7)

Auf die Frage nach Verbesserungsvorschlägen haben die Befragungspersonen eine Reihe von Anregungen gegeben, die im folgenden wörtlich wiedergegeben werden.

Version A: Was könnte besser sein?

- „Im oberen Teil zu voluminös, im unteren Teil zu wenig Erläuterungen.“
- „Alle Felder, die ausgefüllt werden müssen, farblich abgrenzen.“
- „Einrahmung nicht besonders gut.“
- „Differenziertere Fragen zu den einzelnen Punkten.“
- „Klare, gezielte Fragen wären besser.“
- „Die Fragen müßten praxisbezogener sein.“

Version B: Was könnte besser sein?

- „Ziel nicht genau ersichtlich, muß mehr differenziert werden.“
- „Insgesamt gibt es zu wenig Informationen, was alles möglich sein soll und was genau auszufüllen ist. Fragen und gewünschte Ergebnisse passen nicht. Beispiele zielen auf gesamten Betrieb ab und hier wird nur ein Produktionsmonat gewünscht.“
- „Alles! Klare Unterteilung, ob 1., 2., 3. Schicht, nicht ‚übrige‘ Schichten. Klären, ob nur für einen Betriebsbereich oder für alle, dann die 1. Frage ‚echte‘ Produktionstage fragen, nicht Fehltage.“
- „Bessere Anleitungen und Erläuterungen zum Ausfüllen (Beispiel auf der Vorderseite besser); Erläuterungen, z. B. was SYPRO-Nr. ist; wie ist es mit Gleitzeiten (Krankheitstage oder freitags mit 6 Stunden gearbeitet?).“
- „Bei ‚2. Berechnung ...‘: ‚übrige Schichten‘ – muß dies einzeln oder zusammengefaßt aufgeführt werden?“
- „Genauere Erklärungen bei den Fragen, z. B. ob es für jede Abteilung extra oder insgesamt oder nur für eine Abteilung ...“
- „Klare Struktur/Linie fehlt, schon der Kopf des Bogens ist sehr unübersichtlich – wo soll man anfangen?“
- „Die Bereiche, die auszufüllen sind, hervorheben.“

Version C: Was könnte besser sein?

- „Erläuterungen ausführlicher.“
- „Beispiel und Erläuterungen und eigentlicher Fragebogen sollten auf einer Seite sein.“
- „Erläuterung bei Einzelbetrieben fehlt. Erläuterung weist nicht auf verschiedene Werke bzw. Tagesrhythmen hin, nicht besonders gut. Bogen ist nur für Schichtbetriebe geeignet, die mehr als eine Schicht fahren.“
- „Kommentare zu den geforderten Angaben sind zu kurz; müssen konkretisiert werden, z. B. Berichtsmonat. Hinweis, daß die Felder auszufüllen sind, fehlt.“

- „Erläuterungen auf der gleichen Seite wären ideal.“
- „Weniger geheimnisvolle Abkürzungen.“
- „Konkretere Fragen.“

6 Fazit: Empfehlung aus methodischer Perspektive

Bevor wir zu einer zusammenfassenden Bewertung der vorliegenden Ergebnisse kommen, soll zunächst angemerkt werden, daß es überaus hilfreich war, sich für die letztendlich gewählte Form eines Beobachtungspretests mit nachfolgender Befragung der Auskunftspersonen zu entscheiden. Hätte man statt diesem Verfahren eine rein (Daten) erfassende Form der Voruntersuchung gewählt, hätte man also nur die Informationen erfassen und auswerten können, die direkt und unmittelbar auf den zu testenden Bogen enthalten waren bzw. durch sie gewonnen worden wären, wäre es nicht gelungen, Kriterien für eine Entscheidungsfindung zugunsten einer der drei Versionen zu finden: Alle Bogen der Versionen A, B und C wurden weitgehend vollständig ausgefüllt. Aufgrund der formalen Vollständigkeit der ausgefüllten Bogen ist eine Entscheidung zugunsten einer der drei zu testenden Versionen also nicht zu treffen gewesen.

Welche Ergebnisse bringt nun der Beobachtungspretest?

Generell läßt sich sagen, daß aus der Sicht der Auskunftspersonen, wenn man sie direkt danach fragt, ...

... Version A als relativ einfach auszufüllen und als übersichtlich bewertet wird, aber auch als eher uninteressant,

... Version B zwar als schwieriger auszufüllen und etwas weniger übersichtlich gilt als A und C, dafür aber noch am ehesten interessant erscheint und

... Version C ebenfalls einfach auszufüllen ist und einigermaßen übersichtlich erscheint, aber nur wenig interessant.

Sowohl bei Version A als auch bei Version B bezieht sich die deutliche Kritik an der Frage nach den Produktionstagen auf die negative Formulierung der Fragestellung („... keine Produktion...“), die offensichtlich der innerbetrieblichen Denkweise widerspricht. Statt nach Tagen zu fragen, an denen keine Produktion stattfindet, wäre es besser, nach Tagen zu fragen, an denen produziert wird.

Für Version A spricht über das Gesagte hinaus, daß sie insgesamt diejenige Version mit der höchsten Verständlichkeit ist, sowohl gemessen an der direkten Beurteilung durch die Befragten als auch indirekt (über den Aufwand, der mit dem Ausfüllen verbunden ist). Gegen Version A spricht, daß diese Version unter bestimmten betrieblichen Konstellationen (ständig wechselnde Schichten, zu viele Produktionsabteilungen) nicht differenziert genug ist, um die betriebliche Realität abzubilden.

Für Version B spricht allenfalls, daß sie aus der Sicht der Auskunftspersonen immer noch relativ einfach zu bearbeiten ist. Gegen Version B spricht, daß sie den Auskunftspersonen sowohl bei Produktionstagen als auch bei Betriebszeiten weniger verständlich erscheint als

die Versionen A und C. Wie bei Version A zeigt sich auch bei Version B, daß sie betriebliche Realitäten nur undifferenziert und von daher nicht ausreichend abbilden kann.

Für Version C spricht, daß gerade sie genau dieses Defizit von Version A und Version B am besten ausgleichen kann; sie ist am ehesten geeignet, die Komplexität der betrieblichen Realität abzubilden. Probleme mit wechselnden Schichten, zu vielen Produktionsabteilungen etc., wie sie bei den Versionen A und B vorkommen, treffen bei Version C nicht zu. Für Version C spricht auch, daß sie diejenige Version ist, die von den meisten Auskunftspersonen in der kürzesten Zeit zu bearbeiten war (obwohl sich die Unterschiede zu den Versionen A und B relativieren, wenn man eine Ausfülldauer von bis zu zehn Minuten für zumutbar hält).

Sollte Version C beibehalten werden, müßte deutlicher gemacht werden, auf welchen Zeitraum sich die Spalte „Tage“ bezieht; der Begriff „Betriebsteil“ müßte besser definiert und das Schema insgesamt besser erklärt werden. Interessant erscheint auch der Vorschlag einer Auskunftsperson, die Struktur der Tabelle zu ändern: „Schicht 1 – Personal“, „Schicht 2 – Personal“ etc. Das Beispiel zu Version C wird insgesamt relativ gut bewertet; dennoch wären die Begriffe „Betriebsteil“ und „Schichtdauer“ näher zu definieren.

Zieht man alles Gesagte ins Kalkül, so kommen wir zu der Empfehlung, Version C weiterzuverfolgen; sie hat gegenüber den anderen Versionen den Vorteil, daß sie der Komplexität der betrieblichen Realität aufgrund ihrer differenzierten Tabellenstruktur am nächsten kommt. Da dieser Aspekt für die gesamte Erhebung zentral ist, empfehlen wir also aufgrund der Pretestergebnisse eine Entscheidung zugunsten Version C.

Insbesondere die betriebliche Praxis der Großbetriebe mit unterschiedlichen Produktionseinheiten und Betriebe mit vielschichtigen Produktionsstrukturen findet sich in allen Varianten des Erhebungsbogens aber nur kaum oder überhaupt nicht wieder.

In allen Versionen bedarf es daher eines verstärkten Aufwandes bei der praxisorientierten Präzisierung zentraler Begriffe, jeder Bogen sollte gründlich überarbeitet werden.

7 Umsetzung der Pretestergebnisse in einen überarbeiteten Erhebungsbogen

Nach dem Pretest wurde die Variante C des Erhebungsbogens überarbeitet, da sie den Betrieben am einfachsten und den Betriebsverhältnissen am nächsten erschien sowie den geringsten Zeitaufwand verursachte. Im Pretest stellte sich heraus, daß nur ca. 35 % der Betriebe das Feld für die Betriebsnummer ausgefüllt hatten und ca. 50 % der Betriebe ihre SYPRO-Nummer nicht kannten. Für die Haupterhebung wurde daher vorgeschlagen, den Betrieben beide Nummern zur Auswahl vorzugeben, um damit eine korrekte Zuordnung des Erhebungsmerkmals zu den Wirtschaftszweigen zu ermöglichen.

Vielen Betrieben war unklar, für welchen Zeitraum die Tage anzuschreiben waren. Es wurde daher die Kopfspalte „Tage“ durch „Tage im Berichtsmonat“ ergänzt. Der Begriff „Betriebs-
teile“ wurde von den Betrieben als praxisfremd empfunden, so daß er bei der Überarbeitung durch den Begriff „Teil des Produktionsbereichs“ ersetzt wurde.

Um das Verständnis des Erhebungsbogens zu fördern, wurden die Erläuterungen zu den Begriffen, die in der Kopfzeile verwendet wurden, überarbeitet. Im einzelnen wurde zuerst der Begriff „Produktionsbereich“ besser abgegrenzt und durch zusätzliche Beispiele für betriebliche Bereiche, die nicht zu dem Produktionsbereich gehören, ergänzt.

In die Erläuterungen zu dem Begriff „Schichtpersonal“ wurde nach der Definition auf Sonderfälle wie z. B. Teilzeitbeschäftigte oder Leiharbeitskräfte eingegangen. In die Erläuterungen zu den Merkmalen „Schichtdauer“ und „Tage im Berichtsmonat“ wurden Sonderfälle aufgenommen, um eine bessere Abgrenzung der Merkmale zu erreichen. Das Ende der Erläuterungen bildet ein Beispiel für einen Betrieb, bei dem die zuvor genannten Besonderheiten berücksichtigt wurden.

Überarbeitete Version des Erhebungsbogens – Vorderseite

Statistisches Landesamt		Erfassung von Arbeitszeiten im Verarbeitenden Gewerbe	
Unsere Telefon-Nr.:		Betriebs-Nr.:	
		Bei Rückfragen bitte angeben.	
		Bitte senden Sie 1 Exemplar bis zum 5. des folgenden Monats zurück.	
Bitte nur die schattierten Felder ausfüllen!		Erläuterungen, Beispiel und Rechts-	
Name und Anschrift des Betriebes:		grundlagen finden Sie auf der Rückseite.	
		Name des Bearbeiters, für Rückfragen des Statistischen Landesamtes:	
		Telefon-Nr. (für Rückfragen):	
Betriebs-Nr.	SYPRO-Nr.:	Berichtsmonat:	

Beachten Sie bitte die Erläuterungen und das Beispiel auf der Rückseite!

[illegible]

Überarbeitete Version des Erhebungsbogens – Rückseite

Erläuterungen

- Die folgenden Angaben beziehen sich auf den Produktionsbereich eines Werkes "XY". Dazu gehören z. B. auch die Vor-/Nachproduktion und Bereiche, die Hilfsprodukte herstellen. Die Bereiche wie "Rechnungslegung", "Personalwesen", "EDV" und "Lager" z. B. gehören nicht dazu. Alle Angaben beziehen sich auf einen Kalendermonat. Falls der Produktionsbereich eines Werkes "XY" aus mehreren organisatorisch selbständigen Einheiten besteht, können hierfür getrennte Angaben unter der jeweiligen laufenden Nummer gemacht werden.
- Zu dem **Schichtpersonal** gehören alle Arbeitnehmer, die im Produktionsbereich eines Werkes "XY" eingesetzt werden. Die Angabe der Personalstärke bezieht sich dabei auf die jeweilige Schicht. Erhöht sich z. B. eine Schichtstärke durch Leasing-Beschäftigte um mehr als 10 % oder gehen mehr als 10 % der Arbeitnehmer einer Schicht in ein verlängertes Wochenende, so ist dies entsprechend dem unten aufgeführten Beispiel zu vermerken. Teilzeitkräfte können entsprechend ihrem Anteil an der tariflichen Arbeitszeit z. B. als Halbtagskraft mit dem Wert "0,5" angegeben werden.
- Zur **Schichtdauer** zählt die Anwesenheit der Arbeitnehmer an ihrem Arbeitsplatz. Überstunden sind z.B. der betreffenden Schichtdauer hinzuzurechnen, während Kurzarbeit z. B. abzuziehen ist. Die betriebsübliche Mittagspause gehört jedoch nicht zur Schichtdauer. Bei täglichen Veränderungen aus produktionstechnischen Gründen genügt dagegen die Angabe von Durchschnittswerten!
- Es sind für einen Kalendermonat die Tage anzugeben, an denen in den einzelnen Schichten mit dem entsprechenden Personal produziert worden ist. Fällt dagegen in einem Werk "XY" die Produktion an einem Sonntag bzw. an einem Feiertag oder aufgrund eines Maschinen-defektes für einen ganzen Tag aus, dann ist dieser nicht aufzuführen!

- Ein Beispiel für die Eintragung der Personal-, Zeit- und Tagesangaben:

Im Berichtsmonat April ergeben sich für ein Werk "XY" folgende Daten:

- Schicht 1: 35 Arbeitnehmer über 8 Stunden an 20 Tagen,
- Schicht 2: 30,5 Arbeitnehmer über 9 Stunden an 18 Tagen und
- Schicht 2: 26 Arbeitnehmer über 7 Stunden an 2 Tagen (Urlaub für mehr als 10 % der Arbeitnehmer sowie Verkürzung der 2. Schicht aus produktionstechnischen Gründen).

Die o. g. Personal- und Zeitangaben beziehen sich auf den Produktionsbereich eines Werkes (lfd. Nr. 1) und führen zu folgenden Eintragungen:

Teil des Produktionsbereichs	Schichtpersonal				Schichtdauer				Tage im Berichtsmonat	
	Lfd. Nr.	Schicht 1	Schicht 2	Schicht 3	Schicht 4	Schicht 1	Schicht 2	Schicht 3	Schicht 4	Anzahl
1	35				8					20
1		30,5				9				18
1		26				7				2

Zweck, Art und Umfang der Erhebung

Die Erhebung erfolgt im Auftrag des Bundesministers für Wirtschaft. Sie erstreckt sich auf höchstens 10 000 Betriebe im Verarbeitenden Gewerbe, die von der amtlichen Produktionsstatistik erfasst werden. Sie soll Informationen zur Berechnung der Anzahl der monatlichen Produktionstage und der Betriebszeit bereitstellen. Dies dient einer besseren Erfassung konjunktureller Einflüsse sowie einer besseren Berechnung von Produktivitätskennzahlen.

Rechtsgrundlage

§ 7 Abs. 1 des Gesetzes über die Statistik für Bundeszwecke (Bundesstatistikgesetz - BStG) vom 22. Januar 1987 (BGBl. I S. 462, 505), geändert durch Artikel 2 des Gesetzes vom 17. Dezember 1990 (BGBl. I S. 2837). Die Auskunftserteilung ist freiwillig.

Geheimhaltung

Die erhobenen Einzelangaben werden nach § 16 BStG ausnahmslos geheimgehalten. Sie dienen ausschließlich statistischen Zwecken.

Hilfsmerkmale, Löschungen und Ordnungsnummern

Name und Telefonnummer der für Rückfragen zur Verfügung stehenden Personen sind Hilfsmerkmale, die lediglich der technischen Durchführung der Erhebung dienen. Sie werden zusammen mit dem Erhebungsvordruck nach Abschluss der Prüfung der Angaben auf Vollständigkeit und Schlüssigkeit vernichtet. Die Betriebsnummer dient der rationalen Aufbereitung der Ergebnisse und der Identifikation des Betriebes. Mit Hilfe der SYPRO-Nummer kann der Betrieb einem Wirtschaftszweig entsprechend dem Schwerpunkt seiner Tätigkeit zugeordnet werden.

Teilnehmerverzeichnis

A

Akkerboom, Dr. Hans; *Centraal Bureau voor de Statistiek, Heerlen, Niederlande*

B

Beckenbach, Andreas; *Deutsches Krebsforschungszentrum Heidelberg*

Becker, Horst; *polis, Gesellschaft für Politik- und Sozialforschung mbH, München*

Berke, Paul; *Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen, Düsseldorf*

Bick, Dr. Wolfgang; *Amt für Statistik, Wahlen und Einwohnerwesen, Frankfurt am Main*

Blanke, Karen; *Statistisches Bundesamt, Wiesbaden*

Brünnecke, Karin; *Institut für Medizinische Soziologie, Marburg*

C

Chlumsky, Jürgen; *Statistisches Bundesamt, Wiesbaden*

Christochowitz, Sylvia; *Kriminologisches Forschungsinstitut Niedersachsen e. V., Hannover*

D

Dehue, Francine; *Centraal Bureau voor de Statistiek, Heerlen, Niederlande*

Deutscher, Arite; *Statistisches Landesamt des Freistaates Sachsen, Kamenz*

Dobroschke, Wolhard; *Amt für Statistik, Wahlen und Einwohnerwesen, Frankfurt am Main*

Donsbach, Prof. Dr. Wolfgang; *Technische Universität Dresden*

Dorka, Nicolai; *RMM – Dr. Müller Markt- und Sozialforschung GmbH, Hamburg*

Dresch, Dr. Alfred; *Statistisches Bundesamt, Wiesbaden*

Drews, Hans-Peter; *SAMPLE Institut GmbH, Mölln*

E

Ehling, Dr. Manfred; *Statistisches Bundesamt, Wiesbaden*

F

Förster, Hubert; *MMA Media Markt Analysen GmbH, Frankfurt am Main*

Franke, Christine; *IRES GmbH, Düsseldorf*

G

Glade, Katja; *GFM-GETAS GmbH, Hamburg*

Gruber, Winfried; *Statistisches Landesamt Baden-Württemberg, Stuttgart*

Guttmann, Edeltraud; *Statistisches Bundesamt, Zweigstelle Berlin*

H

Hartmann, Dr. Peter H.; *Universität zu Köln*

Hartmann, Dr. Petra; *Zentralarchiv für Empirische Sozialforschung an der Universität zu Köln*

Haupt, Dr. Klaus; *psyma, arbeitsgruppe für psychologische marktanalysen GmbH, Rückersdorf*

Heidenreich, Hans-Joachim; *Statistisches Bundesamt, Wiesbaden*

Helmcke, Thomas; *Statistisches Bundesamt, Wiesbaden*

von der Heyde, Christian; *Infratest Burke AG, München*

Hin, Monika; *Statistisches Bundesamt, Wiesbaden*

Hübbers, Dr. Birgit; *Statistisches Landesamt Rheinland-Pfalz, Bad Ems*

Hüfken, Volker; *Institut für Soziologie, Dresden*

Hüpper, Bernadette; *Deutsches Krebsforschungszentrum Heidelberg*

J

Jäger, Marianne; *Statistisches Bundesamt, Wiesbaden*

Jansen, Rolf; *Bundesinstitut für Berufsbildung, Berlin*

K

Kieslich, Carolin; *GfK Dataservices GmbH, Nürnberg*

Klass, Johannes; *LINK + Partner GmbH, Frankfurt am Main*

Klimesch, Holger; *IM Leipzig Institut für Marktforschung, Leipzig*

Koch, Martina; *Compagnon Marktforschungsinstitut GmbH & Co. KG, Stuttgart*

Koch-Arzberger, Dr. Claudia; *Hessisches Statistisches Landesamt, Wiesbaden*

König, Christian; *Statistisches Bundesamt, Wiesbaden*

Koltzk, Ingo; *Landesamt für Datenverarbeitung und Statistik Brandenburg, Außenstelle Frankfurt/Oder*

Kopsch, Günter; *Statistisches Bundesamt, Wiesbaden*

Kyi, Gregor; *Statistisches Bundesamt, Wiesbaden*

L

Langenhoff, Georg; *Sozialforschungsstelle Dortmund, Außenstelle Eving*

Luiten, Annemiek; *Centraal Bureau voor de Statistiek, Heerlen, Niederlande*

M

Mecklenburg, Eberhard; *Kriminologisches Forschungsinstitut Niedersachsen e. V., Hannover*

Mochmann, Ekkehard; *Zentralarchiv für Empirische Sozialforschung an der Universität zu Köln*

Modest-Wiegand, Angelika; *BASISRESEARCH GmbH, Frankfurt am Main*

Mohler, Prof. Dr. Peter Ph.; *Zentrum für Umfragen, Methoden und Analysen (ZUMA), Mannheim*

Müller, Prof. Dr. Dr. Ulrich; *Institut für Medizinische Soziologie, Marburg*

Münnich, Dr. Margot; *Statistisches Bundesamt, Zweigstelle Berlin*

N

Naujoks, Christel; *Statistisches Bundesamt, Wiesbaden*

Noelle-Neumann, Prof. Dr. Dr. h. c. Elisabeth; *Institut für Demoskopie Allensbach*

Nowack, Regina; *MAFO-Institut Dr. E. Bruckert GmbH & Co. KG, Schwalbach*

O

Ohlemacher, Dr. Thomas; *Kriminologisches Forschungsinstitut Niedersachsen e. V., Hannover*

P

Peiffer, Dietmar; *Infratest Burke AG, München*

Pelmer, Anke; *BASISRESEARCH GmbH, Frankfurt am Main*

Petersen, Thomas; *Institut für Demoskopie Allensbach*

Porst, Rolf; *Zentrum für Umfragen, Methoden und Analysen (ZUMA), Mannheim*

Prüfer, Peter; *Zentrum für Umfragen, Methoden und Analysen (ZUMA), Mannheim*

R

Reim, Uwe; *Statistisches Bundesamt, Wiesbaden*

Rexroth, Margrit; *Zentrum für Umfragen, Methoden und Analysen (ZUMA), Mannheim*

Riede, Thomas; *Statistisches Bundesamt, Wiesbaden*

S

Sander, Ingo; *EMNID-Institut GmbH & Co., Bielefeld*

Schatz, Roland; *Institut für Demoskopie Allensbach*

Scheuch, Prof. Dr. Erwin K.; *Universität zu Köln*

Schneid, Michael; *Zentrum für Umfragen, Methoden und Analysen (ZUMA), Mannheim*

Schraudt, Eva; *IFAK-Institut GmbH & Co., Taunusstein*

Schreiber, Dirk; *Statistisches Bundesamt, Wiesbaden*

Schrey, Ernst; *Statistisches Bundesamt, Wiesbaden*

Schüler, Dr. Klaus; *Statistisches Bundesamt, Wiesbaden*

Schütz, Dieter; *Statistisches Landesamt Saarland, Saarbrücken*

Schwamb, Hans Joachim; *Statistisches Bundesamt, Wiesbaden*

Schwarz, Norbert; *Statistisches Bundesamt, Wiesbaden*

Schwarz, Rainer; *USUMA GmbH, Berlin*

Steffen, Robert; *Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen, Düsseldorf*

Stegemann, Hagen; *Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e. V., Bonn*
Steinkamp, Horst; *Statistisches Landesamt Bremen*
Strecke, Prof. Dr. Heinrich; *Universitäten Tübingen und München, Starnberg*
Stumpf, Klaus; *Marplan Forschungsgesellschaft mbH, Offenbach*

T

Thoma, Michaela; *Zentrum für Umfragen, Methoden und Analysen (ZUMA), Mannheim*

V

Vogelmann, Martin; *Czaia Marktforschung GmbH, Bremen*

W

Walter, Mario; *Bayerisches Landesamt für Statistik und Datenverarbeitung, München*
Wänke, Dr. Michaela; *Universität Heidelberg*
Weichert, Dieter; *Intermarket GmbH, Düsseldorf*
Wein, Elmar; *Statistisches Bundesamt, Wiesbaden*
Wiegand, Erich; *Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V.,
Offenbach*
Wiegert, Dr. Rolf; *Universität Tübingen*

Schriftenreihe Spektrum Bundesstatistik

herausgegeben vom Statistischen Bundesamt

Band 1

Internationaler Workshop 1992 zur Umgestaltung der Agrarstatistik

in den Staaten Mittel- und Osteuropas

169 Seiten · DM 18,80 · Bestellnummer: 1030501-93900 · ISBN 3-8246-0343-8

Beiträge zu einem Workshop des Instituts für Agrarsektoranalyse der Universität Bonn und des Statistischen Bundesamtes zur Umgestaltung der Agrarstatistik in den Staaten Mittel- und Osteuropas in Berlin 1992.

Band 2

Sabine Köhler

Bewertung des Waldes im Rahmen der gesamtwirtschaftlichen Vermögensrechnung

– Möglichkeiten und Grenzen –

82 Seiten · DM 13,80 · Bestellnummer: 1030502-94900 · ISBN 3-8246-0344-6

In diesem Band wird untersucht, inwieweit eine Bewertung des Waldes (Waldboden und Baumbestände) auf der Basis vorhandener Statistiken im Rahmen der gesamtwirtschaftlichen Vermögensrechnung möglich ist und welche statistischen Lücken bestehen. Ausgehend von einer Darstellung der Methoden und Probleme der Waldbewertung in der forstwirtschaftlichen Theorie und Praxis sowie eine Analyse möglicher statistischer Quellen werden Schätzmodelle entwickelt und Größenordnungen aufgezeigt.

Band 3

Verbesserung der statistischen Erfassung der Unternehmensgewinne zur Berechnung des Bruttosozialproduktes von der Einkommenseite

– In Vorbereitung. –

Band 4

Finanzstatistische Kennzahlen für den Bildungsbereich

– In Vorbereitung. –

Band 5

Joachim Klaus u.a.

Umweltökonomische Berichterstattung

– Ziele, Problemstellungen und praktische Lösungsansätze –

244 Seiten · DM 24,80 · Bestellnummer: 1030505-94900 · ISBN 3-8246-0358-6

Diese Studie, die im Rahmen des Aufbaus der Umweltökonomischen Gesamtrechnungen (UGR) in Auftrag gegeben wurde, stellt ausführlich die vielfältigen Ansätze für eine Umweltökonomische Berichterstattung auf nationaler und internationaler Ebene dar – auch unter dem Blickpunkt des nachhaltigen Wirtschaftens. Sie zeigt deutlich die bestehenden Schwierigkeiten und die Bandbreite der Möglichkeiten, an die Probleme heranzugehen, und bildet damit eine Grundlage für die Arbeiten zur Weiterentwicklung der UGR.

**Zu beziehen durch
den Buchhandel oder
den Verlag Metzler - Poeschel
70182 Stuttgart**

**Verlagsauslieferung
Hermann Leins GmbH & Co. KG
Postfach 11 52
72125 Kusterdingen**

Schriftenreihe Spektrum Bundesstatistik

herausgegeben vom Statistischen Bundesamt

Band 6

Michael Reidenbach

Öffentliche Umweltschutzausgaben als Teil der Maßnahmen im Umweltschutz

118 Seiten · DM 15,80 · Bestellnummer: 1030506-94900 · ISBN 3-8246-0378-0

Ergebnisse eines Forschungsprojektes, das im Rahmen der Umweltökonomischen Gesamtrechnungen (UGR) vom Deutschen Institut für Urbanistik (DIFU), Michael Reidenbach, bearbeitet wurde. Im Mittelpunkt steht die Erfassung der öffentlichen Umweltschutzausgaben der verschiedenen föderalen Ebenen der Bundesrepublik Deutschland. Um eine umfassende Information im Sinne der UGR zu ermöglichen, war es notwendig, ein gegenüber vorhandenen Methoden erweitertes und modifiziertes Konzept der Datensammlung zu entwickeln und empirisch anzuwenden.

Band 7

Effizienzbemessung der Hochschulausbildung auf statistischer Grundlage

85 Seiten · DM 14,80 · Bestellnummer: 1030507-95900 · ISBN 3-8246-0460-4

Die aktuellen Bemühungen um eine Reform des deutschen Hochschulsystems zielen vor allem auf eine Erhöhung der Effektivität der Lehre und einen verstärkten Wettbewerb zwischen den Hochschulen. So sollen Lehre und Studium mit Hilfe von vergleichenden hochschulstatistischen Indikatoren transparenter gestaltet und damit ein Beitrag zu einer effizienteren Hochschulausbildung geleistet werden. Der vorliegende Band – mit Beiträgen aus Hochschulplanung, -politik und -statistik – untersucht die Möglichkeiten und Schwierigkeiten bei der Berechnung von Leistungsindikatoren für den Hochschulbereich insbesondere auf der Datengrundlage der amtlichen Statistik.

Band 8

Der Preisindex für die Lebenshaltung aller privaten Haushalte in Gestalt eines Kettenindex

– Beurteilung aus praktischer, empirischer und theoretischer Sicht –

– In Vorbereitung. –

Band 9

Pretest und Weiterentwicklung von Fragebogen

119 Seiten · DM 17,80 · Bestellnummer: 1030507-96900 · ISBN 3-8246-0472-8

Pretestverfahren gewinnen als Instrument der Qualitätsverbesserung sowohl in der Umfrageforschung als auch in der Umfragepraxis zunehmend an Bedeutung. Sie werden zur Vorbereitung von Erhebungen nicht nur von Marktforschungs- und sozialwissenschaftlichen Instituten, sondern auch in der amtlichen Statistik eingesetzt.

Schwerpunkt dieses Bandes sind die einzelnen Beiträge der wissenschaftlichen Tagung „Pretest und Weiterentwicklung von Fragebogen“, die im Juni 1995 im Statistischen Bundesamt stattfand. Die Beiträge befassen sich einerseits mit Methoden zum Test von Erhebungsinstrumenten und andererseits mit der praktischen Umsetzung von Pretests. Ergänzt werden die Tagungsbeiträge um die Erfahrungen aus zwei Pretests, die das Statistische Bundesamt in Zusammenarbeit mit dem Zentrum für Umfragen, Methoden und Analysen (ZUMA), Mannheim, durchgeführt hat.

**Zu beziehen durch
den Buchhandel oder
den Verlag Metzler - Poeschel
70182 Stuttgart**

**Verlagsauslieferung
Hermann Leins GmbH & Co. KG
Postfach 11 52
72125 Kusterdingen**
