

Zur Validität von Fragen in standardisierten Erhebungen: Ergebnisse des Einsatzes eines kognitiven Pretestinterviews

Kurz, Karin; Prüfer, Peter; Rexroth, Margrit

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Kurz, K., Prüfer, P., & Rexroth, M. (1999). Zur Validität von Fragen in standardisierten Erhebungen: Ergebnisse des Einsatzes eines kognitiven Pretestinterviews. *ZUMA Nachrichten*, 23(44), 83-107. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-208195>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

ZUR VALIDITÄT VON FRAGEN IN STANDARDISIERTEN ERHEBUNGEN. ERGEBNISSE DES EINSATZES EINES KOGNITIVEN PRETESTINTERVIEWS¹

KARIN KURZ, PETER PRÜFER UND MARGRIT REXROTH

In diesem Beitrag wird nach einem kurzen Überblick über unterschiedliche Formen des Pretests gezeigt, wie kognitive Pretestverfahren in der Praxis eingesetzt werden können. Dazu werden Beispiele aus dem kognitiven Pretest zur Vorbereitung des ALLBUS 1998 herangezogen, bei dem primär mit Probing-Fragen gearbeitet wurde. Anhand der Beispiele wird demonstriert, daß diese Verfahren tatsächlich aufschlußreiche Hinweise auf das Frageverständnis erbringen können. Gleichzeitig wird deutlich, daß die Anwendung der Verfahren nicht ganz frei von Problemen ist und daß noch Bedarf an methodologischer Forschung zur Konstruktion von Probing-Fragen besteht.

A brief overview of different forms of pretests is presented, followed by demonstrations of how cognitive pretests can be applied in practice using examples from the German General Social Survey (ALLBUS) 1998 cognitive pretest. Probe questions were a prominent feature of this pretest. Examples are used to demonstrate how these techniques can provide valuable information on how respondents understand survey questions. Furthermore, problems associated with using these techniques are considered. It becomes obvious that further research on the construction of probe questions is needed.

1. Einleitung

Zu den zentralen Bedingungen des erfolgreichen Einsatzes eines standardisierten Fragebogens gehört es, daß die Fragen von den Befragten so verstanden werden, wie es vom

¹ Für wertvolle Hinweise zu früheren Fassungen des Beitrages danken wir Achim Koch, Frauke Kreuter sowie zwei anonymen Gutachtern.

Forscher beabsichtigt ist. Wichtige Instrumente zur Überprüfung des Frageverständnisses sind im Rahmen von sogenannten kognitiven Pretestverfahren entwickelt worden. Bei diesen Verfahren wird im Gegensatz zu den gängigen Beobachtungspretests mit kognitiven oder qualitativen Interviewtechniken gearbeitet – also etwa offenen Fragen danach, wie die Befragten zu ihren Antworten gekommen sind (vgl. Kap. 2).

Seit den 80er Jahren werden kognitive Pretests in der internationalen Methodenliteratur diskutiert. Trotzdem konnten sie sich in der Bundesrepublik in den regelmäßig durchgeführten sozialwissenschaftlichen Umfragen nicht durchsetzen. Weder beim Sozioökonomischen Panel noch beim ALLBUS oder dem Wohlfahrtssurvey werden kognitive Pretestverfahren standardmäßig genutzt. Woher rührt diese Zurückhaltung, die sicherlich nicht nur die eben genannten Umfragen betrifft, sondern wohl die große Mehrheit der quantitativen Befragungen, die in sozialwissenschaftlichen Forschungsprojekten durchgeführt wird? Ein wichtiger Grund liegt sicherlich in der Skepsis gegenüber dem Nutzen kognitiver Pretests. Bezeichnenderweise werden auch in neueren deutschen Methodenlehrbüchern kognitive Pretestverfahren entweder überhaupt nicht oder nur am Rande erwähnt (vgl. z.B. Diekmann 1995; Kromrey 1991; Laatz 1993)². Einen allgemein akzeptierten Kanon von kognitiven Pretestverfahren gibt es nicht³ und – zumindest in deutscher Sprache – keine Dokumentationen, wie ein solcher Pretest konkret konstruiert sein sollte. Ein nicht zu unterschätzendes Hemmnis dürfte auch darin liegen, daß für die Durchführung eines kognitiven Pretests ein ganz anderes Fachwissen notwendig ist, sowohl bei der Interviewdurchführung wie auch bei der Auswertung der Ergebnisse, als beim gewohnten Beobachtungspretest – ein Know-How, das von inhaltlich orientierten quantitativen Forschern oftmals erst erworben werden muß. Und nicht zuletzt ist an die Zeit- und monetären Kosten zu denken, denn ein kognitiver Pretest stellt einen erheblichen zusätzlichen Aufwand neben dem Beobachtungspretest („Standardpretest“) dar.

Dennoch, die verbreitete Zurückhaltung gegenüber kognitiven Pretests muß verwundern, liegen doch mittlerweile seit mehr als 15 Jahren (fast ausschließlich englischsprachig) publizierte Erfahrungen mit kognitiven Pretestverfahren vor. Die diversen Monographien und Tagungsbände sprechen eine klare Sprache: kognitive Pretestverfahren – wie Think-Aloud-, Paraphrasing- und Probing-Verfahren – können wesentlich dazu beitragen, Validitätsprobleme von Fragen aufzudecken (vgl. z.B. Belson 1981, 1986; Tanur 1992; Schwarz/Sudman 1996). Vor allem in der Entwicklungsphase von Surveyfragen erweisen

² Eine Ausnahme stellt die neueste Ausgabe des Lehrbuches von Schnell/Hill/Esser (1999) dar, in der erstmals auch kognitive Pretestverfahren skizziert werden.

³ Einen deutschsprachigen Überblick zu kognitiven Pretestverfahren geben Prüfer/Rexroth (1996).

sich kognitive Pretestverfahren als nützlich, während Beobachtungspretests primär als abschließende Tests des Gesamtfragebogens geeignet sind.

In diesem Beitrag möchten wir anhand von Beispielen zeigen, wie ein solcher Einsatz von kognitiven Interviewtechniken konkret aussehen kann. Dabei wollen wir zum einen demonstrieren, daß ein kognitiver Pretest in der Tat aufschlußreiche Hinweise auf das Frageverständnis erbringen kann. Zum anderen möchten wir anhand der Beispiele zeigen, daß die Anwendung in methodischer Hinsicht nicht ganz frei von Problemen ist. Im einzelnen gehen wir wie folgt vor: Im ersten Schritt skizzieren wir den „klassischen“ Beobachtungspretest sowie kognitive Pretesttechniken. Danach geben wir einen Überblick über den beim ALLBUS 1998 eingesetzten kognitiven Pretest. Im dritten Schritt illustrieren wir anhand von vier Beispielen Vorgehensweise, Ergebnisse sowie methodische Probleme. Der Beitrag schließt mit einer Zusammenfassung und Diskussion.

2. Formen des Pretests

Ein Pretest dient der Überprüfung eines Fragebogens und seiner Fragen im Hinblick auf eine Reihe von Aspekten. Dazu gehören – in Anlehnung an Converse/Presser (1986):

- Verständlichkeit der Fragen
- Probleme der Befragten mit ihrer Aufgabe
- Interesse und Aufmerksamkeit der Befragten bei einzelnen Fragen
- Interesse und Aufmerksamkeit der Befragten während des gesamten Interviews
- Wohlbefinden der Befragten (respondent well-being)
- Häufigkeitsverteilung der Antworten
- Reihenfolge der Fragen/unerwünschte Kontexteffekte
- Probleme der Interviewer mit Fragebogen/Befragungshilfen
- Zeitdauer der Befragung

Unter diesen Punkten kommt dem Frageverständnis und den Schwierigkeiten beim Beantworten eine zentrale Bedeutung zu. Nur Fragen, die von den Befragten so verstanden werden, wie vom Forscher beabsichtigt, und zu deren Beantwortung die Befragten in der Lage sind, können zu validen Messungen führen.

In der Praxis (in Deutschland) werden meist Beobachtungspretests („Standardpretests“) angewendet, d.h. es wird beobachtet, welche Probleme und Auffälligkeiten während des

Pretestinterviews auftreten⁴. Diese können z.B. darin bestehen, daß der/die Befragte bei einer Frage um Klärung bittet oder durch Anmerkungen Unsicherheit bei der Antwort erkennen läßt. Die Auffälligkeiten werden in der Regel vom Interviewer nach dem Interview in einem schriftlichen Bericht festgehalten. Der Pretest findet meist unter Feldbedingungen statt, aber mit kleiner Stichprobe, die meist nicht über eine reine Zufallsauswahl, sondern über ein Quotenverfahren gewonnen wurde.

Eine implizite Annahme des Beobachtungspretests ist, daß Verständnisprobleme an den spontanen Reaktionen der Befragten deutlich werden. Dies ist aber nur selten der Fall, wie Belson (1981, 1986) in umfangreichen Untersuchungen nachgewiesen hat. Fragen werden häufig formal korrekt beantwortet, ohne irgendwelche Anzeichen von Schwierigkeiten oder Mißverständnissen seitens der Befragten, und trotzdem liegt ein falsches - oder besser: nicht beabsichtigtes - Frageverständnis zugrunde. Dieses Problem ist seit langem bekannt. Schon in den 60er Jahren schlug Schuman (1966) vor, bei einzelnen (zufällig ausgewählten) Fragen in der Haupterhebung gezielte Nachfragen zum Frageverständnis zu stellen („random probes“) und damit mehr Klarheit über die Validität von Fragen zu bekommen. Und auch schon in den 40er und 50er Jahren gab es vereinzelt Versuche, durch Nachfragen („probes“) mehr über das Frageverständnis zu erfahren (Cantril 1944; Nuckols 1953).

Wichtige Anstöße brachten Anfang der 80er Jahre mehrere gemeinsame Konferenzen von kognitiven Psychologen und Umfrageforschern in den USA. Im Anschluß daran wurden kognitive Forschungslabors eingerichtet, die sich speziell mit der Anwendung von Konzepten der kognitiven Psychologie beschäftigten, um den Verlauf des Frage-Antwort-Prozesses zu untersuchen und die Tauglichkeit von Fragebögen zu überprüfen (Tanur 1992: ix-xiii). Dabei kam dem US-amerikanischen National Center for Health Statistics eine Vorreiterrolle zu (vgl. Fowler 1995: 111). In der Bundesrepublik waren entsprechende Forschungsaktivitäten vor allem bei ZUMA angesiedelt (vgl. Hippler/Schwarz/Sudman 1987).

Als gemeinsame Kennzeichen der vielfältigen Anwendungsformen kognitiver Interview-techniken im Pretest sind folgende zu nennen (vgl. Fowler 1995: 111; Forsyth/Lessler 1991):

1. Primäres Ziel ist es herauszufinden, wie Befragte eine Frage verstehen und wie sie zu ihrer Antwort kommen. Es werden keine besonderen Anstrengungen unternommen,

⁴ Die Befragten sind dabei nicht über den Testcharakter der Befragung informiert. Converse/Presser (1986) führen für diese Vorgehensweise den Begriff „undeclared pretest“ ein.

die Datenerhebungsprozeduren, die in der Haupterhebung zur Anwendung kommen, zu replizieren.

2. Die Interviews finden oftmals in einem besonderen Setting statt, in dem sie aufgezeichnet und beobachtet werden können. Deshalb wird häufig auch der Begriff „Laborinterview“ verwendet.
3. Anders als im Beobachtungspreset sind die Interviewer in aller Regel keine regulären Feldinterviewer, sondern kognitive Psychologen, Mitglieder des Forschungsteams oder speziell geschulte Interviewer. Die Interviewer kennen meist die Forschungsziele gut und sind damit in der Lage, Diskrepanzen zwischen dem, was die Forscher beabsichtigten und dem, was der/die Befragte tut, zu erkennen.
4. Die Vorgehensweise besteht in Grundzügen immer darin, daß die Fragen zunächst vorgelesen und von dem/der Befragten beantwortet werden; danach (nach der einzelnen Frage oder nach dem gesamten Interview) wird mit Hilfe verschiedener Techniken herauszufinden versucht, was in den Befragten während des Frage-Antwort-Prozesses vorgegangen ist⁵.
5. Die Stichprobe besteht normalerweise aus wenigen (fünf bis zehn), eher unsystematisch gewonnenen Befragten. Besonders wichtig ist, diesen Personen zu vermitteln, daß sie keine herkömmliche Befragtenrolle übernehmen, sondern als „Mithelfende“ einen wichtigen Beitrag zur Verbesserung des Fragebogens leisten sollen. Dies ist auch der Hauptgrund dafür, den Befragten für diese mitunter nicht ganz einfache Aufgabe ein Honorar zu zahlen. Um eine gleichbleibend hohe Konzentration und Motivation seitens der Befragten sicherzustellen, sollte die Dauer der Befragung im allgemeinen eine bis eineinhalb Stunden nicht wesentlich überschreiten. Dies führt in aller Regel dazu, daß nur ein Teil aller Fragen der Haupterhebung getestet werden kann.

Im folgenden sollen die wichtigsten kognitiven Verfahren kurz vorgestellt werden.

Think-Aloud

Der/die Befragte wird aufgefordert, „laut zu denken“ und dabei sämtliche Gedankengänge, die zur Antwort führen bzw. führten, zu formulieren. Ziel dabei ist, aus den Äußerungen der Befragten Hinweise darüber zu erhalten, wie die ganze Frage oder einzelne Begriffe verstanden wurden. Bei der Anwendung der Think-Aloud-Methode gibt es zwei Vorgehensweisen:

⁵ Eine Ausnahme von dieser Vorgehensweise bildet die Concurrent-Think-Aloud-Technik: Hier soll der/die Befragte sofort nach dem Vorlesen der Frage und noch vor der Antwort seine/ihre Gedankengänge verbalisieren.

1. Die Befragten werden aufgefordert, laut zu denken, während sie ihre Antwort formulieren. Diese Vorgehensweise bezeichnet man als Concurrent-Think-Aloud-Methode.
2. Die Befragten werden aufgefordert, nach der Beantwortung der Frage darüber nachzudenken, wie die Antwort zustande kam. Diese Vorgehensweise ist bekannt unter dem Begriff Retrospective-Think-Aloud-Methode.

Paraphrasing

Befragte sollen - nach der Beantwortung - die Frage mit eigenen Worten wiederholen.

Confidence Rating

Befragte sollen nach der eigentlichen Antwort den Grad der Verlässlichkeit ihrer Antwort bewerten (meist mit Hilfe einer Skala, wie z.B.: „Wie sicher waren Sie sich bei Ihrer Antwort: sehr sicher, eher sicher, eher unsicher, sehr unsicher?“).

Probing

Beim Probing handelt es sich um eine altbekannte Interview-Technik, die z.B. zentraler Bestandteil des bereits genannten Verfahrens „random probe“ von Schuman (1966) und des „intensive interview“ von Belson (1981) ist. Dabei wird eine gegebene Antwort vom Interviewer durch eine oder mehrere Zusatzfragen (Probes) „hinterfragt“. Gängige Probing-Varianten sind (vgl. z.B. Oksenberg/Cannell/Kalton 1991):

- Comprehension Probing: Probing zum Frageverständnis bzw. zu speziellen Aspekten der Frage (z.B. Verständnis einzelner Begriffe).
- Information Retrieval Probing: Probing zu Aspekten der Informationsbeschaffung. Sinnvolles Anwendungsgebiet sind besonders retrospektive Faktfragen.
- Category Selection Probing: Nachfragen zu dem gewählten Skalenwert.

Neben den Arbeiten aus dem Bereich der kognitiven Psychologie gaben die Untersuchungen von Belson (1981, 1986) wichtige Anstöße zu einer stärkeren Beschäftigung mit Problemen des Frageverständnisses. Belson entwickelte als Pretestinstrument ein sogenanntes „intensive interview“, um die Validität von Surveyfragen zu untersuchen. In einer verkürzten Form besteht ein solches Interview aus zwei Teilen (vg. Belson 1981: 390ff.). Zunächst wird das Interview wie in einem Standardpretest durchgeführt. Danach folgt der zweite Interviewteil, in dem jeweils drei bis vier vorher festgelegte Fragen des Fragebogens einer intensiven Überprüfung unterzogen werden. Die Überprüfung besteht aus zwei Schritten: Im ersten Schritt liest der Interviewer die zu testende Frage sowie die aus dem Standardpretest vorliegende Antwort noch einmal vor, und die Befragten sollen genau darlegen, wie sie zu ihrer Antwort gekommen sind. Sind die Ausführungen zu ungenau, wird mit zusätzlichen Probing-Fragen versucht, die Befragten zu motivieren, genauere Angaben zu machen. Im

zweiten Schritt wird durch gezielte Probing-Fragen versucht herauszufinden, wie die in der Frage enthaltenen Konzepte von den Befragten verstanden werden.

Daß der Einsatz kognitiver Techniken im Pretestbereich weit mehr relevante Informationen liefert als die klassische Beobachtungsmethode, steht außer Zweifel. Dabei darf jedoch nicht vergessen werden, daß die verschiedenen kognitiven Techniken zum einen nicht alle das Gleiche leisten (Oksenberg/Cannell/Kalton 1991; Presser/Blair 1994) und zum anderen die Anwendungsprozeduren mit unterschiedlichem Aufwand verbunden sind. Im folgenden soll kurz auf einige Problempunkte bei einzelnen Verfahren hingewiesen werden.

- Die Think-Aloud-Technik – als die zentrale Technik der kognitiven Methoden überhaupt – stellt in der Reihe der kognitiven Techniken die höchsten Anforderungen an die Befragten. Bei der Concurrent-Think-Aloud-Variante sind Befragte häufig nicht fähig, die Prozesse, die zur Entscheidung für ihre Antwort führen, unmittelbar zu artikulieren. Schwierigkeiten zeigen sich auch, wenn Befragte nach der Antwortvergabe ihre Überlegungen für das Zustandekommen ihrer Antwort beschreiben sollen. Tatsächlich nennen Befragte dann eher Begründungen und Rechtfertigungen für ihre Antwort. Auf Grund dieser Schwierigkeiten gibt es Befürworter für die eine oder andere Variante: Sudman/Bradburn/Schwarz (1996) bevorzugen eher die retrospektive Variante, da ihrer Meinung nach die Concurrent-Think-Aloud das intellektuelle Leistungsvermögen „normaler“ Befragter überschreitet. Loftus (1984) setzte dagegen die Concurrent-Think-Aloud-Technik erfolgreich bei der Überprüfung retrospektiver Fragen ein. Bisweilen werden generelle Zweifel geäußert, ob mit Hilfe der Think-Aloud-Technik tatsächlich die kognitiven Prozesse, die beim Beantworten insbesondere unter Feldbedingungen ablaufen, abgebildet werden können (Groves 1996: 397). Darüber hinaus liegt es auf der Hand, daß je nach Fähigkeit zur Introspektion und Verbalisierung Befragte unterschiedlich gut in der Lage sind, diese Technik anzuwenden (Fowler 1995: 112). Belegt ist, daß die Fähigkeit zur Verbalisierung mit dem Bildungsstand der Befragten zunimmt (Bickart/Felcher 1996).
- Die Technik des Paraphrasing, bei der die Befragten nach der Beantwortung der eigentlichen Frage den Fragetext in eigenen Worten wiederholen sollen, ist in der Anwendung für Forscher und Befragte zwar einfach durchführbar, sie eignet sich aber nur bedingt zum Erkennen von Verständnisproblemen. So macht ihre Anwendung bei kurzen Faktfragen wenig Sinn. Ergiebig ist das Verfahren bei Meinungsfragen mit langen oder schwierigen Texten. Weiterhin arbeitet die Technik nicht zuverlässig bei der Überprüfung ganz konkreter Aspekte. Sie kann dem Forscher aber Hinweise geben, ob Fragen überhaupt nicht verstanden wurden. Grundsätzlich ist bei der Anwendung des Verfahrens darauf zu achten, daß die Befragten den Fragetext in

eigenen Worten wiedergeben, denn erst diese Umsetzung läßt eventuelle Verständnisprobleme erkennen.

- Die Confidence Rating Technik ermittelt anhand eines Skalenwertes, als wie verlässlich Befragte ihre Antworten einschätzen. Unter Anwendungs- und Auswertungsgesichtspunkten ist die Technik einfach. Zentraler Nachteil: Das eigentliche Problem, das einer wenig verlässlichen Antwort zugrunde liegt, kann – wenn überhaupt – nur indirekt identifiziert werden.
- Die Probing Technik zielt darauf ab, nach der Beantwortung der eigentlichen Frage durch weitere Zusatzfragen vom Befragten Informationen über das Fragenverständnis und/oder den Antwortprozeß zu bekommen. Generell gilt: Je fokussierter Probing-Fragen auf konkrete Aspekte zielen, desto höher ist auch die Wahrscheinlichkeit, Informationen über den zu überprüfenden Aspekt/Inhalt zu bekommen. Die Problematik liegt in der Formulierung dieser Zusatzfragen. Hier ist an ein Problem zu denken, das jedem quantitativen Forscher bewußt sein dürfte, das aber im Rahmen von qualitativen (kognitiven) Pretests nur selten thematisiert wird (vgl. Groves 1996: 399ff.): Frageformulierungen beeinflussen die gegebenen Antworten. Mit Groves (1996: 401) ist daran zu erinnern, daß Instrumente, die die Qualität von Messungen bestimmen sollen, Instrumente größerer Validität erfordern. Doch zur Frage der Validität und auch der Reliabilität von Probing-Fragen gibt es bislang wenig systematische Forschung. William Foddy (1996) setzt sich kritisch mit der Konstruktion von Probing-Fragen auseinander und demonstriert anhand von Beispielen, daß schlecht konstruierte Probing-Fragen nicht effektiv arbeiten können.

3. Kognitiver Pretest in der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS)

Kognitive Pretests sind vor allem bei der Entwicklung von Fragen sinnvoll, Beobachtungspretests hingegen für das abschließende Testen des Gesamtfragebogens. Dementsprechend setzten wir in einer frühen Phase der Fragebogenentwicklung für den ALLBUS 1998 einen kognitiven Pretest ein. Zum Abschluß wurde dann ein Beobachtungspretest durchgeführt. Im Zentrum des kognitiven Pretests standen Fragen, von denen wir vermuteten, daß sie durch vage Begriffe, Fremdwörter, komplizierten Satzbau oder eine nicht alltägliche Perspektive möglicherweise Verständnisprobleme hervorrufen würden. Im Anschluß an Belson (1981) gingen wir von der Hypothese aus, daß die Befragten ihnen Unverständliches oder vage Formuliertes (meist stillschweigend) in Richtung des Gewohnten abändern würden. Unser Hauptanliegen bestand deshalb primär in der Auf-

deckung des Frageverständnisses, das den Antworten zugrundelag und weniger in der detaillierten Abbildung der kognitiven Prozesse, die zu den Antworten führten.

Entsprechend unseres Erkenntnisziels wählten wir die kognitiven Verfahren aus: Da es uns weniger um das Verständnis der gedanklichen Prozesse per se ging, verzichteten wir – bis auf eine Ausnahme (Schätzung der Fernsehnutzungsdauer) – auf die Think-Aloud-Technik. Ein Argument gegen dieses Verfahren waren auch die hohen Anforderungen, die es an die Befragten stellt. Primär verwendeten wir Probing-Techniken: Bei Fachbegriffen und vagen Ausdrücken stellten wir direkte Nachfragen zum Begriffsverständnis („comprehension probing“). Sollten mehrere Begriffe von den Befragten zusammenfassend bewertet werden, fragten wir, inwieweit diese Bewertung Schwierigkeiten bereitete. Bei komplexen Aussagen oder Fragen mit nicht alltäglicher Perspektive ermittelten wir, was den Befragten bei der Beantwortung durch den Kopf gegangen sei oder warum sie eine bestimmte Antwort gegeben hatten. Mit diesen Probing sollte herausgefunden werden, wie die Frage aufgefaßt worden war und inwieweit Abweichungen vom intendierten Frageverständnis aufgetreten waren. Wenn wir konkrete Hypothesen über mögliche Fehlerquellen hatten, setzten wir zusätzlich spezifische Fragen ein. Bei einer Frage nutzten wir das Paraphrasing-Verfahren (vgl. Abschnitt 4.4).

Wir stellten die Probing-Fragen jeweils direkt im Anschluß an die Beantwortung der für die Hauptuntersuchung geplanten Frage. Dies hat den Vorteil, daß die Überlegungen und Gedankenschritte den Befragten noch eher präsent sein dürften als nach dem Interview (Fowler 1995: 113). Nachteilig ist, daß Kontexteffekte nicht geprüft werden können, da die Surveyfragen durch die Nachfragen unterbrochen werden.

Insgesamt 18 Fragen wurden für den kognitiven Pretest ausgewählt, und zwar einerseits Fragen, die erstmals im ALLBUS 1998 zum Einsatz kommen sollten, andererseits auch Fragen, die als bewährt gelten und schon in früheren ALLBUS-Erhebungen gestellt worden waren. Die Fragen stammten primär aus den Bereichen politische Einstellungen, politische Partizipation sowie Mediennutzung. An dem kognitiven Pretest nahmen 17 nach einem Quotenplan ausgewählte Personen teil⁶, die bis dahin noch keine Pretester-

⁶ Der Quotenplan wurde nicht vollständig erfüllt; ursprünglich waren 20 Interviews geplant; ältere Personen wurden seltener erfaßt. In der realisierten Stichprobe befanden sich acht Frauen, elf Personen mit max. Hauptschulabschluß, vier Personen mit mittlerer Reife, zwei Personen mit Abitur. Sieben Befragte rekrutierten sich aus der Altersgruppe der 18- bis 34jährigen, sieben aus den 35- bis 54jährigen, und drei waren 55 Jahre oder älter. Acht Personen waren Vollzeit berufstätig, drei Teilzeit oder nebenher beschäftigt; sechs Personen waren nicht berufstätig (darunter ein Student, ein Rentner, zwei Arbeitslose und vier Hausfrauen).

fahrung hatten. Hauptschulabsolventen wurden überproportional erfaßt. Im Hintergrund stand die Annahme, daß Befragte mit niedriger Bildung eher Probleme mit dem Fragebogen haben. Deshalb erschien eine höhere Fallzahl für diese Gruppe sinnvoll. Der kognitive Pretest fand zum Teil im „Labor“ statt, d.h. in den Räumen der Feldabteilung bei ZUMA, aber auch im Feld, d.h. bei den Befragten zu Hause. Die Interviews wurden von den Autoren dieses Beitrages sowie zwei erfahrenen Interviewern der Feldabteilung durchgeführt. Die Antworten wurden während der Befragung schriftlich protokolliert; gleichzeitig wurden zur Sicherheit sämtliche Interviews auf Band aufgezeichnet.

4. Beispiele

Die konkrete Formulierung von Probing, inhaltliche Ergebnisse, aber auch einzelne methodische Probleme sollen im folgenden anhand von vier Beispielen veranschaulicht werden. Bei den Beispielen 1, 2 und 4 handelt es sich um Fragen bzw. Items, die zuvor noch nicht im ALLBUS enthalten waren. Beispielfrage 3 war schon 1988 im ALLBUS gestellt worden.

4.1 Nachfrage zur Erläuterung von Begriffen

Beginnen wir mit dem einfachsten Anwendungsfall, in dem geklärt werden sollte, wie die Befragten bestimmte Begriffe in einer Aussage verstehen. Die Technik, um ein solches Begriffsverständnis zu überprüfen, ist denkbar einfach: Wir fragten im Anschluß an die Antwort nach, was der/die Befragte unter den Begriffen verstanden hatte.

Beispiel 1: „Volksbegehren und Volksentscheide sind eine notwendige Ergänzung der repräsentativen Demokratie.“

Die Aussage war Bestandteil einer Itematterie, in der nach dem Grad der Zustimmung zu bestimmten Demokratievorstellungen und Einstellungen zu Demokratienormen gefragt wurde⁷. Wir vermuteten, daß die Begriffe „Volksbegehren“, „Volksentscheid“ sowie „repräsentative Demokratie“ nicht von allen Befragten verstanden würden. Deshalb stellten wir die folgenden Nachfragen:

„Was verstehen Sie unter ‘Volksbegehren’?“

„Was verstehen Sie unter ‘Volksentscheid’?“

„Und was verstehen Sie unter ‘repräsentativer Demokratie’?“

⁷ Die Antwortmöglichkeiten waren „stimme voll und ganz zu“, „stimme eher zu“, „stimme eher nicht zu“ und „stimme überhaupt nicht zu“.

Unsere Vermutung, daß viele Befragten ein mangelndes Verständnis dieser Begriffe haben würden, wurde im großen und ganzen bestätigt. Die Ergebnisse sollen hier nicht im Detail präsentiert werden. Als grober Befund ist festzuhalten, daß etwa die Hälfte der Antworten bei Volksbegehren und Volksentscheid ungefähr in die richtige Richtung gingen. Bei einzelnen dieser Antworten blieben aber einige Zweifel am korrekten Begriffsverständnis bestehen, denn die Umschreibungen lagen häufig sehr nahe an den Begriffen, z.B. Volksentscheid: „eine Entscheidung, die das Volk getroffen hat“ oder Volksbegehren: „Wunsch der Bürger“.

Bei dem Begriff „repräsentative Demokratie“ war die Situation noch eindeutiger: 15 von 17 Befragten wußten keine bzw. nur eine klar falsche Antwort auf die Nachfrage.

Beispiele:

Id. 01: „Die Demokratie wird hier repräsentiert, die wird eben repräsentiert, dargestellt.“

Id. 07: „Daß unterschiedliche Meinungen gelten, diese auch vertreten werden können, daß man Demokratie praktiziert und wirklich dahintersteht, das ist repräsentativ.“

Id. 15: „Schwierig, was soll ich sagen, wenn man sie vorzeigen kann, wenn andere Länder sagen: Da schau mal her!“

Die Ergebnisse, wenngleich aufgrund der Stichprobe nicht zu verallgemeinern, machen sehr deutlich, daß ein großer Teil der Befragten zumindest einen Teil der in der Aussage enthaltenen Begriffe nicht verstand. Wie könnte eine valide Messung des angesprochenen Sachverhalts – direkte Demokratieformen als Ergänzung zur repräsentativen Demokratie – aussehen? Der Zusatz „repräsentativ“ bei dem Wort „Demokratie“ erscheint verzichtbar. Worauf es ankommt, ist allerdings die Benennung von direkten Demokratieformen, und hier scheinen sich „Volksbegehren“ und „Volksentscheid“ nicht zu eignen. Wenngleich ebenfalls nicht optimal, bietet sich der Begriff „Volksabstimmung“ an, der in einem anderen Item enthalten war und ebenfalls getestet wurde. Bei Nachfragen zum Begriff „Volksabstimmung“ gaben immerhin 12 von 17 Befragte Antworten, die in die richtige Richtung gingen. Da das Wort „Volksabstimmung“ allein durch seine Formulierung den tatsächlichen Inhalt nahelegt: „Das Volk stimmt über etwas ab“, ist es wohl für Befragte vergleichsweise einfacher zu verstehen. Das revidierte Item, das so im ALLBUS 1998 realisiert wurde, lautete:

„Volksabstimmungen sind ein notwendiger Bestandteil der Demokratie.“

Trotzdem bleiben auch bei diesem Item Zweifel daran, ob es tatsächlich ein valides Meßinstrument darstellt. Denn auch beim Begriff „Volksabstimmung“ gab es einige Befragte, die den Begriff nicht erläutern konnten oder falsche Angaben machten. Eine bessere

Lösung würde ohne Zweifel darin bestehen, den Befragten vor der Frage zunächst eine Begriffserklärung für das Wort „Volksabstimmung“ zu geben.

Methodische Ergebnisse

Trotz ihrer Einfachheit erwies sich die gewählte Nachfrage – „Was verstehen Sie unter ...?“ – als nicht optimal. Wie schon mehrfach angesprochen, blieben die Befragten mit ihren Antworten oftmals relativ nahe am genannten Wort, so daß zu vermuten ist, daß sie den Begriff lediglich in seine Bestandteile aufgliederten und daraus einen Satz bildeten. Unklar bleibt dabei offensichtlich, ob die Bedeutung des Begriffes tatsächlich bekannt war. Erfolgversprechender wäre daher eine Probing-Frage, bei der ein solches Antwortverhalten von vornherein ausgeschlossen oder zumindest unwahrscheinlicher ist. Als mögliche Nachfrage böte sich z.B. an: „Stellen Sie sich vor, Sie müßten jemanden den Begriff ... erklären, was würden Sie da sagen?“

4.2 Nachfrage zur Begründung der Antwort

Das im folgenden ausgewählte Beispiel erschien auf den ersten Blick relativ unproblematisch. Der Satzbau ist allerdings vergleichsweise komplex, und es wird ein vager Ausdruck verwendet: „etwas erreichen können“.

Beispiel 2: „Ein Bürger sollte sich nur dann politisch engagieren, wenn er glaubt, damit auch etwas erreichen zu können.“

Die Aussage war Bestandteil einer Batterie von Items zu Demokratienormen und Einstellungen zur Demokratie. Die Befragten sollten anhand einer Skala angeben, wie stark sie den einzelnen Aussagen zustimmen⁸. Das mit Beispiel 2 ausgewählte Item thematisiert Kosten-Nutzen-Überlegungen, die bei der Entscheidung über individuelles politisches Engagement möglicherweise angestellt werden. Die Fragenkonstrukteure wollten wissen, wie verbreitet eine „zweckrationale“ Norm individueller politischer Partizipation ist, nach der ein Bürger sich nur dann engagieren sollte, wenn er damit auch sein Ziel erreichen kann.

Nachdem die komplette Fragebatterie gestellt und beantwortet worden war, wurden einzelne Items noch einmal vorgelegt und jeweils die folgende Nachfrage gestellt.

„Warum haben Sie hier [Skalenwert einsetzen] genannt?“

⁸ Die Antwortmöglichkeiten waren „stimme voll und ganz zu“, „stimme eher zu“, „stimme eher nicht zu“ und „stimme überhaupt nicht zu“.

Aus den Antworterläuterungen läßt sich erschließen, ob die Aussage in der beabsichtigten Weise verstanden wurde. Bei sieben der 17 Befragten geht aus der Antwort hervor, daß sie die Frage verstanden hatten. Die Ansichten dieser Befragten waren geteilt: ein Teil der Befragten war der Meinung, daß man sich nur dann politisch engagieren sollte, wenn man glaubt, etwas erreichen zu können, der andere Teil meinte, daß man sich dann engagieren soll, wenn man hinter einer Sache steht, auch wenn das gesteckte Ziel nicht unmittelbar erreicht wird. Eine Person gab als Begründung für ihre mangelnde Zustimmung, daß man es vorher nicht einschätzen könne, ob man tatsächlich etwas erreicht.

Beispiele:

- Id. 01* („stimme eher zu“): „Weil es sinnlos ist, wenn man sich politisch engagiert und nichts erreichen kann.“
- Id. 02* („stimme voll und ganz zu“): „Sonst wäre es Kraftverschwendung.“
- Id. 04* („stimme überhaupt nicht zu“): „Man sollte nie immer nur tun, was man erreichen kann, vor allem, wenn man hinter einer Sache steht.“
- Id. 05* („stimme überhaupt nicht zu“): „Bürger sollte immer politische Meinung haben und sich einsetzen...“

Vier Personen interpretierten das Item nicht ganz so, wie es beabsichtigt war:

- Id. 12* („stimme eher nicht zu“): „Jeder sollte sich immer verantwortlich fühlen, engagieren, auch wenn man keinen persönlichen Vorteil daraus zieht.“ [„Etwas erreichen können“, wird negativ im Sinne eines persönlichen Vorteils interpretiert.]
- Id. 16* („stimme überhaupt nicht zu“): „Denke, auch einzelner Bürger kann etwas bewirken, wenn er glaubt, was wichtig ist und er im kleinen anfängt, und das Kleine hilft schon.“ [„Etwas erreichen können“ wird im Sinne von „etwas Großes erreichen können“ interpretiert.]
- Id. 10* („stimme eher zu“): „Wie soll ich das begründen... wenn jemand nichts macht, kann er nichts erreichen. Wenn er glaubt, daß er was erreichen kann, muß er dies machen.“ [Es bleibt unklar, ob man sich nur dann engagieren sollte, wenn man glaubt etwas erreichen zu können.]
- Id. 9* („stimme voll und ganz zu“): „Wenn einer von Politik keine Ahnung hat, kann er mit seinem Geschwätz nichts erreichen. Er macht dann mehr kaputt als ganz.“ [Ausdruck „glaubt, etwas erreichen zu können“ wurde nicht beachtet.]

Eine Person lehnte es ab, daß man sich überhaupt politisch engagieren sollte und stimmte der Aussage deshalb nicht zu. Und schließlich mißverstanden zwei Personen das Item ganz, und zwei Personen konnten keine Begründung nennen.

Insgesamt erweckten die Antworten auf die Nachfrage den Eindruck, daß die Mehrheit der Befragten (11 Personen) die Frage ungefähr in der beabsichtigten Weise verstanden hatte. Allerdings interpretierten, wie erwähnt, immerhin vier dieser Befragten die Aussage

nicht ganz wie erwartet. Woraus die Fehlinterpretationen resultierten, läßt sich bei mehreren Personen aus den Antworten erschließen: Erstens scheinen kleine, aber bedeutsame Elemente der Aussage überhört zu werden, wodurch die Gesamtaussage falsch interpretiert wird. Dies gilt für die Worte „nur“ und „glaubt“ (etwas erreichen zu können). Zweitens läßt der Aussageteil „etwas erreichen zu können“ einen weiten Interpretationsspielraum zu und führt damit zu unvorhergesehenen Interpretationen.

Die Hinweise aus den Antworten wurden genutzt, um die Verständlichkeit des Items zu erhöhen. Das Wort „nur“ wurde an prominente Stelle – an den Anfang des Satzes – gestellt, so daß es weniger leicht übersehen oder überhört wird. Der Ausdruck „etwas erreichen“ wurde durch „etwas bewirken“ ersetzt, da dieser Ausdruck einen weniger breiten Interpretationsspielraum zuläßt. „Etwas bewirken“ läßt sich nicht im Sinne des Erreichens eines persönlichen Vorteils deuten. Allerdings bleibt die Unbestimmtheit des Ausdruckes „etwas“ natürlich bestehen.

Das neu formulierte und so im ALLBUS 1998 realisierte Item lautete deshalb:

„Nur wenn man glaubt, etwas bewirken zu können, sollte man sich politisch engagieren.“

Methodische Ergebnisse

Die Methode, über die Begründung oder nähere Erläuterung der Antwort Hinweise zum Frageverständnis zu erhalten, erwies sich in diesem Beispiel wie auch bei weiteren Anwendungen als fruchtbar (vgl. auch Beispiel 4). Zumindest klare Mißverständnisse der Frage oder Aussage waren an den Erläuterungen der Antworten ablesbar.

4.3 Frage zum Antwortprozeß und zusätzliche Nachfrage

Die im folgenden getestete Frage wurde bereits mehrfach im ALLBUS und auch in anderen Umfragen eingesetzt; erstmals von Kaase 1985 in der Studie „Ansprüche der Bürger an den Staat“. Mit der Frage soll die Zufriedenheit mit dem politischen System der Demokratie in der Bundesrepublik Deutschland gemessen werden.

Beispiel 3:

Interviewer: Bitte Liste 3 vorlegen!

Kommen wir nun zu der Demokratie in der Bundesrepublik Deutschland: Wie zufrieden oder unzufrieden sind Sie – alles in allem – mit der Demokratie, so wie sie in der Bundesrepublik Deutschland besteht?

- | | |
|--------------------|--------------------------|
| sehr zufrieden | <input type="checkbox"/> |
| ziemlich zufrieden | <input type="checkbox"/> |
| etwas zufrieden | <input type="checkbox"/> |

etwas unzufrieden	<input type="checkbox"/>
ziemlich unzufrieden	<input type="checkbox"/>
sehr unzufrieden	<input type="checkbox"/>
weiß nicht	<input type="checkbox"/>

Überprüft werden sollte, inwieweit diese Frage tatsächlich – wie beabsichtigt – im Sinne von Zufriedenheit mit dem System (diffuse politische Unterstützung) beantwortet wird oder ob nicht die Zufriedenheit mit der aktuellen Politik der Regierung (spezifische Unterstützung), die mit einer weiteren (nachfolgenden) Frage gemessen wurde, in das Zufriedenheitsurteil mit einfließt.

Im Unterschied zum vorherigen Beispiel baten wir die Befragten zunächst nicht um eine Begründung ihrer Antwort, sondern zielten mit einer ersten Nachfrage direkt auf die Wiedergabe des Antwortprozesses:

1. „Als Sie eben die Frage beantwortet haben, an was haben Sie da gedacht? Was ist Ihnen alles durch den Kopf gegangen, bis Sie die Antwort gegeben haben?“

Danach fragten wir konkret nach, was das Zufriedenheitsurteil bestimmte:

2. „Was macht Sie ...[Antwort wiederholen]?“ (Falls bei der Beantwortung der Frage „weiß nicht“: „Warum haben Sie „weiß nicht“ geantwortet?“)

Methodische Ergebnisse

Während der Interviews wurde deutlich, daß die zweite Nachfrage unter methodischen Gesichtspunkten wenig sinnvoll war. Es hat den Anschein, daß die zweite Nachfrage dazu anregte, frei Dinge anzuführen, die für die tatsächliche Antwort weniger relevant waren, und zwar einfach deshalb, weil die für die Antwort relevanten Dinge schon auf die erste Nachfrage hin genannt worden waren. Verstärkend kam hinzu, daß in der Nachfrage nicht mehr explizit auf den gedanklichen Prozeß des Antwortens oder die Gründe für die gegebene Antwort Bezug genommen wurde. Die Probleme wurden von mehreren Interviewern erkannt: In einigen Interviews wurde deshalb auf die zweite Nachfrage verzichtet. In den Interviews, bei denen die zweite Nachfrage gestellt wurde, wichen die Antworten zum Teil erheblich von denen zur ersten Nachfrage ab. Aus diesen Gründen wird im folgenden auf die inhaltliche Interpretation der Antworten zur zweiten Nachfrage verzichtet.

Inhaltliche Ergebnisse⁹

Alle Befragten wählten eine der vorgegebenen Antwortalternativen aus. Einer der Befragten verbalisierte in seiner spontanen Nachfrage genau das oben beschriebene Problem: „Ist nur Demokratie gemeint oder auch Regierung?“

Bei der ersten Nachfrage bezogen sich knapp die Hälfte – nämlich acht Befragte – auf Systemmerkmale der Demokratie (Wahlrecht, Meinungsfreiheit etc.) und zum Teil auf den Vergleich mit anderen Staaten. Es ist demnach anzunehmen, daß sie die Frage zumindest ungefähr in der beabsichtigten Weise verstanden.

Beispiele:

Id. 03 ('sehr zufrieden'): „Intuitiv geantwortet, denke an meine vielen Freiheiten, meine Rechte und habe das mit anderen Ländern verglichen, und das schätze ich sehr.“

Id. 14 ('ziemlich unzufrieden'): „Daß eigentlich der Bürger im großen und ganzen fast nirgends mitentscheiden kann.“

Id. 15 ('ziemlich zufrieden'): „Wir dürfen wählen, wir können frei entscheiden.“

Id. 13 ('ziemlich unzufrieden'): „Für mich bedeutet Demokratie direkte Demokratie. Wir haben Parteiendemokratie. Momentan ist kein großer Unterschied zwischen den etablierten Parteien, deshalb ist für mich Demokratie fragwürdig. Demokratie bei denen = Kapitalisierung.“

Dagegen nannten vier Personen aktuelle politische Probleme wie 'Arbeitslosigkeit' oder 'Ausländer'; eine Person bezog sich auf ihre persönliche Zufriedenheit.

Beispiele:

Id. 08 ('ziemlich unzufrieden') : „Erstens mein Gehaltszettel und die Vorstellung an meine zukünftige Rente, auch mit Ausländerzufluß in Deutschland, mit Arbeitslosigkeit bei Jugendlichen, fehlende Perspektive in vielen Bereichen.“

Id. 10 ('ziemlich zufrieden'): „Arbeitslosigkeit, soziale Absicherung, außenpolitische Sachen.“

Bei diesen Befragten hat es den Anschein, daß sie eher den Grad ihrer Zufriedenheit mit der augenblicklichen wirtschaftlichen und politischen Situation bzw. der Regierung ausdrückten. Ein Befragter nannte sowohl Systemmerkmale wie auch spezifische politische Probleme. Die übrigen drei Antworten auf die Nachfragen waren so allgemein, daß sie keinen Rückschluß auf das im Sinne des Forschers 'richtige' oder 'falsche' Frageverständnis zulassen.

⁹ Die hier vorgestellten Ergebnisse nutzen die Befunde aus einer ersten Auswertung von Martina Wasmer, ZUMA, Mannheim.

Inhaltlich läßt sich somit festhalten, daß es Probleme bei der Messung des abstrakten Konzeptes „Systemzufriedenheit“ gibt. Die Probleme mögen zum Teil darauf zurückzuführen sein, daß über die Institution „Demokratie“ zu urteilen, eine Aufgabe ist, die eher wenig alltäglich ist. Im Alltag präsenter sind sicherlich konkrete politische Probleme, die mit dem Handeln von Politikern in der Demokratie in Verbindung gebracht werden. Durch die Formulierung im Nachsatz – „... Demokratie, so wie sie in der Bundesrepublik Deutschland besteht“ – wird zudem der Blick auf die konkrete Situation gelenkt.

Was könnte unternommen werden, um das Verständnis der Befragten in die beabsichtigte Richtung zu lenken? Zu vermuten ist, daß durch eine andere Formulierung – etwa: „... mit der Demokratie in der Bundesrepublik Deutschland“ – die Aufmerksamkeit weniger stark auf tagespolitische Probleme gerichtet würde. Weiterhin könnten durch einen Einleitungstext bereits vor den beiden Fragen zur diffusen und spezifischen politischen Unterstützung die Dimensionen klargestellt werden. Etwa so: „In den beiden folgenden Fragen geht es zum einen um Ihre Zufriedenheit mit der Demokratie in Deutschland im allgemeinen und zum anderen um Ihre Zufriedenheit mit den gegenwärtigen Leistungen der Bundesregierung. Kommen wir zunächst ...“. Für den ALLBUS kamen allerdings derartige Veränderungen nicht in Frage, da durch Umformulierungen die Vergleichbarkeit über die Zeit und mit anderen Ländern gefährdet worden wäre.

4.4 Paraphrasing kombiniert mit Nachfragen

Mit der im folgenden behandelten Frage sollte erfaßt werden, für wie wirksam die Befragten ihre *persönliche* Beteiligung an verschiedenen politischen Partizipationsformen halten. Die Einschätzung der Wirksamkeit des persönlichen Handelns wird im Rahmen eines Rational Choice Ansatzes von entscheidender Bedeutung für die Nutzenerwartung und damit die Wahrscheinlichkeit einer Handlung angesehen. Der hier aufgeführten Frage gingen Fragen zur tatsächlichen Beteiligung an den genannten (sowie einigen weiteren) politischen Partizipationsmöglichkeiten voraus.

Als generelles Problem vermuteten wir die *Nichtalltäglichkeit* der Fragestellung. Personen bewerten, so unsere Erwartung, politische Partizipationsformen im Alltag eher im Hinblick auf ihre *generelle* Wirksamkeit als im Hinblick auf die Wirksamkeit ihrer *persönlichen* Beteiligung. Davon ausgehend war unsere Hypothese, daß zumindest einige Befragte die Frage stillschweigend in Richtung des Gewohnten oder Alltäglichen abändern (vgl. auch Belson 1981: 371ff.)

Beispiel 4:

Interviewer: Kärtchenspiel vorlegen!

Kommen wir noch einmal zu den unterschiedlichen politischen Aktivitäten. Für wie wirksam halten Sie Ihre persönliche Beteiligung an diesen Aktivitäten, um ein politisches Anliegen zu erreichen? Bitte sagen Sie mir jeweils, ob Sie Ihre persönliche Beteiligung als sehr wirksam, ziemlich wirksam, nicht sehr wirksam oder gar nicht wirksam einschätzen.

	sehr wirksam	ziemlich wirksam	nicht sehr wirksam	gar nicht wirksam	weiß nicht
A Seine Meinung sagen, im Bekanntenkreis und am Arbeitsplatz	<input type="checkbox"/>				
B Sich an Wahlen beteiligen	<input type="checkbox"/>				
D Mitarbeit in einer Bürgerinitiative	<input type="checkbox"/>				
E In einer Partei aktiv mitarbeiten	<input type="checkbox"/>				
H Hausbesetzung, Besetzung von Fabriken, Ämtern	<input type="checkbox"/>				
M Teilnahme an einer genehmigten Demonstration	<input type="checkbox"/>				
N Sich aus Protest nicht an Wahlen beteiligen	<input type="checkbox"/>				

Um das vermutete Problem zu untersuchen, setzten wir drei verschiedene Methoden ein.

„Würden Sie bitte die eigentliche Frage noch einmal mit eigenen Worten wiederholen. Um was geht es bei dieser Frage?“ (Paraphrasing)

Damit sollte überprüft werden, ob die Befragten tatsächlich von persönlicher Wirksamkeit sprechen würden oder ob sie das Wort „persönlich“ mehr oder weniger bewußt ignorieren und entsprechend nicht wiedergeben würden. Um ganz sicher zu gehen, daß die Befragten tatsächlich ihre persönliche Beteiligung im Kopf hatten, stellten wir als zweite Nachfrage:

„Haben Sie bei den genannten Aktivitäten immer an ihre eigene Beteiligung gedacht?“

Und schließlich wollten wir wissen, ob sich die Art, wie die Wirksamkeitseinschätzung vorgenommen wird, zwischen verschiedenen Aktivitäten unterscheidet. Wir fragten bei drei Items (A, B, M) genauer nach:

„Was meinen Sie genau damit, wenn Sie sagen, daß Sie diese Aktivität als [hier genannten Skalenwert einsetzen] einschätzen?“

Methodische Ergebnisse

Die verschiedenen Methoden erwiesen sich als unterschiedlich brauchbar. Auffällig ist zunächst, daß beim Paraphrasing drei Personen angaben, die Frage vergessen zu haben – ein Hinweis darauf, daß sie die Frage nicht verstanden hatten. Problematisch erscheint, daß in einigen Fällen das Paraphrasing deutlich andere Ergebnisse brachte als die anschließenden konkreten Nachfragen. Dabei zeigte sich vor allem, daß der beim Paraphrasing wiedergegebene Frageinhalt nicht unbedingt zu dem später bekundeten Vorgehen bei der Beantwortung der einzelnen Items paßte (3. Nachfrage). Dies kann zwei Ursachen

haben: Zum einen wurde in unserer Paraphrasing-Aufforderung nicht nach dem tatsächlichen Frageverständnis gefragt (sondern nach der Wiedergabe des Inhaltes), und zum anderen legten die Items je nach Inhalt und Formulierung unterschiedliche Vorgehensweisen bei der Wirksamkeitsbeurteilung nahe (vgl. inhaltliche Ergebnisse).

Die Konstruktion der Nachfrage: „Haben Sie immer an ihre eigene Beteiligung gedacht?“ war ebenfalls unzulänglich. Wie bei den nachfolgenden Angaben zu den Items deutlich wurde, konnte es durchaus vorkommen, daß ein Befragter an die eigene Beteiligung dachte, aber eine Beurteilung im Hinblick auf die generelle Wirksamkeit der Beteiligungsform vornahm. D.h., unsere Nachfrage griff zu kurz, denn wir wollten eigentlich herausfinden, ob die Befragten an die Wirksamkeit der eigenen Beteiligung, also die persönliche Wirksamkeit, gedacht hatten. Unsere Nachfrage war also kein valides Instrument zur Messung dieses Aspektes.

Am effektivsten erwies sich die Aufforderung an die Befragten, für einzelne Items auszuführen, was sie mit ihren Antworten genau gemeint hatten. Hier waren keine methodischen Probleme erkennbar.

Inhaltliche Ergebnisse

Wie eben schon angedeutet, erwiesen sich am aufschlußreichsten die Begründungen, die die Befragten für die Wirksamkeitseinstufungen der drei Items gaben. Deshalb konzentrieren wir uns im folgenden auf diese Ergebnisse.

Keiner der Befragten schien die Frage nach der Wirksamkeit der *persönlichen* Beteiligung an diversen politischen Partizipationsformen tatsächlich *durchgängig* mit Bezugnahme auf die eigene Beteiligung zu beantworten. Bei Item A wurden die Befragten vermutlich durch die Formulierung „seine Meinung sagen“ angeregt, an die Aktivität einer einzelnen Person – naheliegenderweise wohl die der eigenen Person – zu denken. Bei insgesamt neun von 17 Befragten geht aus den Antworten hervor, daß sie an die eigene Meinungsäußerung und deren Wirksamkeit gedacht hatten. Bei den anderen bleibt dies unklar, bzw. bei einzelnen hat es den Anschein, daß sie sich generell auf individuelle Meinungsäußerungen im Bekanntenkreis bezogen hatten.

Beispiele für Bezugnahme auf die Wirksamkeit der eigenen Meinungsäußerung:

Id. 04 („nicht sehr wirksam“): „Ich diskutiere nur mit Bekannten, das dringt nicht an irgendwelche wichtige Stellen.“

Id. 06 („ziemlich wirksam“): „Sind meistens jüngere Leute, und ich glaube, sie akzeptieren meine Meinung, oder sie glauben, daß ich recht habe.“

Id. 07 („nicht sehr wirksam“): „Weil in meinem Bekanntenkreis eine ähnliche Meinung herrscht.“

Id. 13 („nicht sehr wirksam“): „Ich erreiche damit nur einen kleinen Kreis von Leuten, die selber keinen Einfluß haben.“

An den zitierten Antworten wird auch deutlich, daß die Befragten unterschiedliche Erfolgskriterien zur Beurteilung der Wirksamkeit der Aktivität anlegten (Einfluß auf die Meinung der eigenen Bekannten vs. Einfluß bei den ‘eigentlich wichtigen Stellen’).

Bei Item B wird zwar ähnlich wie bei A auf die Aktivität von Personen Bezug genommen („sich an Wahlen beteiligen“), aber im Unterschied zu Item A wird eine institutionalisierte Form politischer Beteiligung angesprochen. Es zeigt sich nun, daß drei Befragte mit der Wirksamkeit von Wahlen allgemein argumentierten:

Id. 02 („nicht sehr wirksam“): „Weil immer nur die großen Parteien an die Macht kommen, das Spektrum ist gering.“

Id. 04 („ziemlich wirksam“): „Weil ich denke, daß die Politiker letztendlich doch ihre eigene Meinung durchsetzen und sich durch Wahlen nur bestätigen lassen.“ [Weshalb sind Wahlen dann „ziemlich wirksam“?]

Id. 13 („nicht sehr wirksam“): „Politische Parteien unterscheiden sich nicht sehr voneinander, von daher hat die Wahl geringe Wirksamkeit.“

Eindeutig auf ihre eigene Wahlbeteiligung oder die individuelle Stimme (nicht notwendigerweise die eigene) zielten die Antworten von 11 Befragten. Beispiele:

Id. 01 („ziemlich wirksam“): „Weil jede Stimme zählt, aber weil es nur eine ist, halte ich sie nur für ziemlich wirksam.“

Id. 05 („sehr wirksam“): „Wenn ich gewisse Partei nicht an der Macht haben will und wenn ich nicht wähle, dann gebe ich denen indirekt einen Vorzug, deshalb muß man wählen.“

Id. 12 („ziemlich wirksam“): [spontane Äußerung schon vor der Nachfrage] „Besser: ‘Ich beteilige mich an Wahlen’. Funktioniert jedoch nur auf der allgemeinen Ebene, wenn eigene Ebene, dann nicht sehr wirksam.“

Id. 15 („sehr wirksam“): „Für mich sehr wichtig, daß ich zu Wahlen gehe, egal, ob sich etwas ändert, für mein gutes Gewissen, wenn’s alle machen würden, sähe es etwas anders aus heute.“

Wie einzelne der Beispiele zeigen, bezogen sich aber nicht alle Befragten auf die Wirksamkeit der einzelnen Stimme. Insgesamt drei der 11 Befragten antworteten im Sinne von persönlicher Wichtigkeit oder innerer Verpflichtung, wählen zu gehen (vgl. ID 15). Ebenfalls drei Befragte schienen implizit an die Wirksamkeit aller Stimmen, die in ihrem Sinne abgegeben werden, zu denken (vgl. ID 5). Bei insgesamt drei der 17 Befragten war die Begründung für ihre Wirksamkeitseinstufung nicht unmittelbar verständlich.

Bei Item M („Teilnahme an einer genehmigten Demonstration“) ging anders als bei den beiden anderen Items *keiner* der Befragten auf die Wirksamkeit der *persönlichen* Beteili-

gung ein. Statt dessen bezogen sich fast alle Befragten auf die Wirksamkeit der Partizipationsform 'Demonstration'. Dabei wurde am häufigsten (von sieben Befragten) mit der Aufmerksamkeit, die Demonstrationen in den Massenmedien, der Öffentlichkeit oder bei den Politikern erzeugen, argumentiert, z.B.:

Id. 01 („ziemlich wirksam“): „Die Massenmedien werden aufmerksam.“

Id. 04 („ziemlich wirksam“): „Weil das an die entsprechenden Stellen gelangt durch TV, Presse, dadurch werden Politiker erst aufmerksam.“

Insgesamt hat es damit den Anschein, daß Inhalt und Formulierung der Items ausschlaggebend für die unterschiedlich starke Bezugnahme auf die persönliche Beteiligung sind. Aufgrund des vorgefundenen Musters der Begründungen bei den Items A, B und M ist zu vermuten, daß mit Itemformulierungen, die noch einmal explizit machen, daß es um die persönliche Beteiligung geht, die Befragten eher angeregt werden, an die eigene Beteiligung zu denken (z.B. „seine Meinung sagen...“). Gleichzeitig hat es aber den Anschein, daß mit der Nennung des Begriffes für die jeweilige Beteiligungsform („Wahlen“, „Demonstration“) bei einigen Befragten sofort eine generelle Wirksamkeitseinstufung dieser Beteiligungsform (und nicht der eigenen Beteiligung) vorgenommen wird. Aus letztgenanntem Grund läßt sich unseres Erachtens keine völlig befriedigende Lösung finden, um die persönliche Wirksamkeitseinschätzung für verschiedene Beteiligungsformen zu erfragen. Für die Haupterhebung des ALLBUS 1998 wurde folgende Frageformulierung gewählt:

„Kommen wir noch einmal zu den unterschiedlichen politischen Aktivitäten. Bitte sagen Sie mir, in welchem Maße Sie persönlich auf die Politik Einfluß nehmen könnten, wenn Sie die Handlungen, die auf diesen Karten beschrieben sind, ausführen würden.

Der Wert 1 bedeutet, daß Sie persönlich dadurch überhaupt nicht auf die Politik Einfluß nehmen könnten, der Wert 7 bedeutet, daß Sie persönlich dadurch auf die Politik sehr stark Einfluß nehmen könnten. Mit den Werten dazwischen können Sie abstufen.

Indem ich¹⁰

A ... mich an Wahlen beteilige

G ... mich an einer genehmigten Demonstration beteilige

L ... an einer Verkehrsblockade teilnehme.“

¹⁰ Im folgenden wird nur eine kleine Auswahl aus den insgesamt 11 Items wiedergegeben. Das Item: „Seine Meinung sagen, im Bekanntenkreis und am Arbeitsplatz“ (bzw. eine Umformulierung dieser Aussage) wurde nicht mehr in die Frage aufgenommen.

5. Diskussion

Das Hauptanliegen unseres kognitiven Pretests war zu untersuchen, wie Befragte mit Frageanforderungen umgehen, die besondere Kenntnis von Fachbegriffen, besondere sprachliche Kompetenzen oder die Übernahme einer nicht alltäglichen Perspektive verlangen. Können wir unter diesen Umständen valide Messungen erwarten? Welche Hinweise geben die Befragtenaussagen auf mögliche Änderungen der Frageformulierungen, die die Meßgenauigkeit erhöhen könnten? Wie eingangs ausgeführt gibt es kein einfaches, allgemein akzeptiertes „Rezept“ für kognitive Pretests in Umfragen. Wir entwickelten daher ein Instrument, bei dem hauptsächlich mit allgemeinen und gezielten Fragen zum Frage- und Begriffsverständnis gearbeitet wird.

Insgesamt waren die Ergebnisse des kognitiven Pretests ermutigend: Probleme und ein vom Beabsichtigten abweichendes Frageverständnis waren meist klar erkennbar. Darüber hinaus, und an dieser Stelle teilen wir die Kritik von Groves (1996) an diesen Verfahren nicht, ergaben sich aus den Ausführungen der Befragten oftmals konkrete Hinweise, wie ein Problem gelöst werden kann. Sinnvoll wäre es natürlich, die Neu- oder Umformulierung einer Frage wieder in einem kognitiven Pretest zu prüfen. Eine solche zweite Untersuchung wurde beim ALLBUS 1998 nicht durchgeführt.

Unsere ersten Erfahrungen mit diesen Methoden machten auch auf die unterschiedliche Brauchbarkeit verschiedener Vorgehensweisen bei den Nachfragen aufmerksam. Schon die einfache Nachfrage, was der/die Befragte unter einem Begriff verstehe, konnte nicht uneingeschränkt als valides Instrument eingestuft werden. Dagegen erwiesen sich in unserer Untersuchung die beiden grundlegenden Probing-Techniken zum Erschließen des Frageverständnisses als relativ unproblematisch: einerseits die Bitte nach Begründung der gegebenen Antwort und andererseits die Aufforderung an die Befragten, noch einmal zu sagen, was ihnen während der Antwort durch den Kopf gegangen ist.

Häufig bestehen schon vorab konkrete Vorstellungen darüber, welcher Punkt in einer Frage oder Aussage den Befragten Probleme bereiten bzw. Mißverständnisse hervorrufen könnte. In solchen Fällen erscheint es ratsam, eine spezifische Nachfrage zu diesem Punkt zu stellen. Dabei sollte auf eine vorherige allgemeine (direkte oder indirekte) Nachfrage zum Antwortprozeß aber nicht verzichtet werden, denn nur dadurch läßt sich ein Eindruck von allen – nicht immer vorhersehbaren – Problemen gewinnen. Als Vorgehensweise bietet sich ein zweistufiger Weg an, bei dem zunächst eine allgemeine und dann eine oder mehrere konkrete Nachfragen gestellt werden. Zu warnen ist davor, „blind“ nach einem ersten allgemeinen Probing noch eine weitere relativ allgemeine Nachfrage zu stellen. Nur wenn sich der/die Befragte noch nicht zu dem in der zweiten Nachfrage angesprochenen Punkt geäußert hat, sollte diese gestellt werden. Ansonsten läuft man Gefahr, daß Befragte Dinge

berichten, die während des Antwortens tatsächlich keine Rolle gespielt haben. Generell ist zu konstatieren, daß noch Forschungsbedarf dazu besteht, welche Nachfragen besonders effektiv zur Erfassung von welchen Problemen sind (vgl. Groves/Fultz/Martin 1992).

Als nicht sehr effektiv erwies sich die Technik des Paraphrasing, da sie offensichtlich nur sehr groben Aufschluß über das Frageverständnis geben kann. Deshalb erscheint es ratsam, Paraphrasing nur in Kombination mit allgemeinen oder spezifischen Probingfragen zu verwenden.

Für primär inhaltlich orientierte Forscher mag sich die Frage stellen, ob die Anwendung eines kognitiven Pretest-Interviews tatsächlich lohnt. Gibt es nicht andere und billigere Wege, um zu guten Fragen zu kommen? Reicht es nicht aus, bewährte Checklisten oder „Kunstlehren“ der Frageformulierung zu konsultieren sowie einen Beobachtungspretest durchzuführen, um zu guten, d.h. validen Fragen zu kommen? Die Frage muß klar verneint werden. Checklisten zur Fragekonstruktion sind natürlich hilfreich, wenn sie ernsthaft beachtet werden. Die Erfahrung zeigt allerdings, daß Probleme trotzdem übersehen oder – noch wichtiger – als nicht gravierend erachtet werden. Um die Angemessenheit der eigenen Einschätzung zu überprüfen, ist deshalb die Durchführung eines kognitiven Pretests ausgesprochen lehrreich. Darüber hinaus erbringt ein kognitiver Pretest Erkenntnisse über das tatsächliche Frageverständnis und Antwortverhalten, die durch keine „Kunstlehren“ der Frageformulierung gewonnen werden können. Auch Beobachtungspretests können offensichtlich nicht als Ersatz für einen kognitiven Pretest fungieren. Das Problem von Beobachtungspretests wurde schon eingangs benannt: Wenn Befragte ein abweichendes Frageverständnis oder Antwortschwierigkeiten haben, müssen diese nicht notwendigerweise vom Interviewer beobachtbar sein. Und darüber hinaus: Die Art der Abweichungen und Probleme wird von den Befragten in den seltensten Fällen während des Beobachtungspretests verbalisiert.

Alles in allem erscheint es deshalb ratsam, im Zuge der Entwicklung von Surveyfragen möglichst viele Fragen mit kognitiven Pretesttechniken zu untersuchen, wobei allerdings ein kognitiver Pretest den abschließenden Beobachtungspretest nicht ersetzen kann. Auch Fragen, bei denen man relativ sicher ist, daß sie für die Befragten unproblematisch sind, sollten einem kognitiven Pretest unterzogen werden. Sinnvoll erscheint es schließlich, diese Pretesttechniken innerhalb der Methodenausbildung für SozialwissenschaftlerInnen praktisch anzuwenden. Denn um den eigenen Blick für gute Surveyfragen zu schärfen, sind kognitive Pretests ausgesprochen erhellend.

Korrespondenzadresse

*Dr. Karin Kurz
Universität Bielefeld
Postach 10 01 31
D-33501 Bielefeld*

Literatur

- Belson, W. A., 1981: The design and understanding of survey questions. Aldershot: Gower.
- Belson, W. A., 1986: Validity in survey research. Aldershot: Gower.
- Bickart, B./Felcher, M. E., 1996: Expanding and enhancing the use of verbal protocols in survey research. S. 115-142 in: N. Schwarz/S. Sudman (Hrsg.), Answering questions. San Francisco: Jossey-Bass Publishers.
- Cantril, H., 1944: Gauging public opinion. Princeton, NJ: Princeton University Press.
- Converse, J. M./Presser, S., 1986: Survey questions. Handcrafting the standardized questionnaire. Quantitative Applications in the Social Sciences No. 63. Newbury Park u.a.: Sage.
- DeMaio, T. J./Rothgeb, J. M., 1996: Cognitive interviewing techniques. In the lab and in the field. S. 177-195 in: N. Schwarz/S. Sudman (Hrsg.), Answering questions. San Francisco: Jossey-Bass Publishers.
- Diekmann, A., 1995: Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. Reinbek bei Hamburg: Rowohlt.
- Foddy, W., 1995: Probing: A dangerous practice in social surveys? *Quality & Quantity* 29: 73-86.
- Foddy, W., 1998: An empirical evaluation of in-depth probes used to pretest survey questions. *Sociological Methods & Research* 27: 103-133.
- Forsyth, B. H./Lessler, J. T., 1991: Cognitive laboratory methods: A taxonomy. S. 393-418 in: Biemer, P.N. et al. (Hrsg.), Measurement errors in surveys. New York: John Wiley.
- Fowler, F. J., 1992: How unclear terms affect survey data. *Public Opinion Quarterly* 56: 218-231.
- Fowler, F. J., 1995: Improving survey questions. Design and evaluation. *Applied Social Research Methods Series*. Volume 38. Thousand Oaks u.a.: Sage.
- Groves, R. M./Fultz, N. H./Martin, E., 1992: Direct questioning about comprehension in a survey setting. S. 49-61 in: Tanur, J. M. (Hrsg.), Questions about questions. Inquiries into the cognitive bases of surveys. New York: Russell Sage Foundation.

- Groves, R. M., 1996: How do we know what we think they think is really what they think? S. 389-402 in: Schwarz, N./Sudman, S. (Hrsg.), *Answering questions*. San Francisco: Jossey-Bass Publishers.
- Hippler, H. J./Schwarz, N./Sudman, S., 1987: *Social information processing and survey methodology*. New York: Springer.
- Kromrey, H., 1991: *Empirische Sozialforschung*. 5. Auflage. Opladen: Leske und Budrich.
- Laatz, W., 1993: *Empirische Methoden. Ein Lehrbuch für Sozialwissenschaftler*. Thun: Verlag Harri Deutsch.
- Nuckols, R., 1953: A note on pre-testing public opinion questions. *Journal of Applied Psychology* 37: 119-120.
- Oksenberg, L./Cannell, Ch./Kalton, G., 1991: New strategies for pretesting survey questions. *Journal of Official Statistics*, 7: 349-365.
- Presser, S./Blair, J., 1994: Survey pretesting: Do different methods produce different results? *Sociological Methodology*, 1994.
- Prüfer, P./Rexroth, M., 1996: Verfahren zur Evaluation von Survey-Fragen: Ein Überblick. *ZUMA-Nachrichten* 39: 95-116.
- Schnell, R., 1991: Pretests in standardisierten Interviews. Unveröffentlichtes Manuskript. Universität Mannheim.
- Schnell, R./Hill, P. B./Esser, E., 1999: *Methoden der empirischen Sozialforschung*. 6. Auflage. München: Oldenbourg.
- Schwarz, N./Sudman, S. (Hrsg.), 1996: *Answering questions. Methodology for determining cognitive and communicative processes in survey research*. San-Francisco: Jossey-Bass Publishers.
- Schuman, H., 1966: The random probe: A technique for evaluating the validity of closed questions. *American Sociological Review*, 31: 218-222.
- Tanur, J. M. (Hrsg.), 1992: *Questions about questions. Inquiries into the cognitive bases of surveys*. New York: Russell Sage Foundation.