

Simone Tries, M. A., Sabine Sattelberger, M. A.

Qualitätssicherung bei qualitativen Pretests

Grundlage für eine fundierte Fragebogenoptimierung

Seit einigen Jahren führt das Statistische Bundesamt Fragebogentests für Haushalts- und Unternehmensstatistiken durch. Möglich wurde dies durch die Einrichtung eines hausinternen Pretestlabors mit Kameras, Mikrofonen und Eye Tracker in den Jahren 2007 und 2008. Mithilfe von Testpersonen evaluiert der Pretestbereich sowohl Papier- als auch Online-Erhebungsinstrumente, die von der Fachstatistik anhand der gegebenen Empfehlungen überarbeitet werden. Durch leicht auszufüllende Formulare soll zum einen die Belastung der Befragten reduziert, zum anderen die Qualität amtlicher Daten erhöht werden. Bei Pretests im Statistischen Bundesamt werden hauptsächlich qualitative Testmethoden eingesetzt.

Einheitliche Standards für die Analyse qualitativer Daten liegen in der Sozialforschung kaum vor; sofern vorhanden, bedürfen diese insgesamt der praktischen Weiterentwicklung. Infolgedessen werden Ergebnisse, die auf qualitativen Methoden beruhen, oft angezweifelt: Sie seien nicht zuverlässig, nicht repräsentativ und nicht signifikant.¹ Um solcher Kritik entgegenzuwirken, arbeitet der Pretestbereich des Statistischen Bundesamtes beständig an der Weiterentwicklung eigener Qualitätsstandards für die Durchführung und Analyse qualitativer Pretests.

Der vorliegende Beitrag² schildert, welche Standards derzeit zur Qualitätssicherung bei qualitativen Pretests in der

amtlichen Statistik umgesetzt werden. Dazu wird zunächst aufgezeigt, wie im Statistischen Bundesamt Testbefragungen für Papier- und Online-Fragebogen typischerweise ablaufen und warum dabei vor allem qualitative Testmethoden zum Einsatz kommen. Die gängigsten Qualitätskriterien der qualitativen Sozialforschung werden dann auf einzelne Prozessschritte angewandt und mit einigen Praxisbeispielen veranschaulicht. Der Beitrag schließt mit einem Ausblick auf künftige Entwicklungen in der Qualitätssicherung des Pretestbereichs.

1 Qualitative Pretests beim Statistischen Bundesamt

1.1 Institutioneller Hintergrund

Alle Statistischen Ämter der Mitgliedstaaten der Europäischen Union und das Statistische Amt der Europäischen Gemeinschaften (Eurostat) haben sich auf die Einhaltung des Verhaltenskodex Europäische Statistiken ("Code of Practice") verpflichtet, der 15 gemeinsame Qualitätsstandards für den gesamten Statistikprozess festlegt. Unter Grundsatz 8 heißt es: „Die Fragebogen für statistische Erhebungen werden vor der Erhebung der Daten systematisch erprobt.“³ Praktische Empfehlungen für die Umsetzung dieser Vorgabe gibt Eurostat im "Handbook of Recommended Practices for Questionnaire Development and Testing in the

¹ Siehe Miles, M. B./Huberman, A. M.: "Qualitative Data Analysis: An Expanded Sourcebook", zweite Auflage, Thousand Oaks 1994, hier: Seite 262 f.

² Dieser Beitrag beruht auf einem englischsprachigen Paper, das die Autorinnen auf der European Conference on Quality in Official Statistics in Helsinki präsentiert haben (siehe Sattelberger, S./Tries, S.: "Quality of Pretesting: Instruments for Evaluation and Standardization", im Internet unter http://q2010.stat.fj/media/presentations/session-23/sattelberger_tries_q2010_paper_qualityofpretesting_paper.pdf, abgerufen am 11. Mai 2011). Ergänzt wurden grundlegende Weiterentwicklungen, welche die Durchführung und Analyse von Tests für Online-Fragebogen und für Fragebogen von Unternehmensstatistiken betreffen.

³ Siehe Statistisches Bundesamt: „Europäische Statistiken Verhaltenskodex“ (im Internet unter www.destatis.de, Pfad: Publikationen → Qualitätsberichte, abgerufen am 11. Mai 2011).

European Statistical System”⁴. Nach dieser Publikation ist der Einsatz qualitativer Pretestmethoden ein probates Mittel, um Ursachen für Antwortausfälle und fehlerhafte Angaben in Fragebogen aufzudecken. Es fehlen jedoch konkrete Leitlinien zur Auswertung qualitativer Pretestdaten, vor allem weil statistische Ämter traditionell eher mit quantitativen Daten arbeiten. Folglich ist die Entwicklung eigener Qualitätsstandards bezogen auf qualitative Daten eine notwendige Neuerung in der deutschen amtlichen Statistik.⁵

1.2 Methodische Gründe für qualitative Pretests

Grundsätzlich lassen sich zwei Hauptkategorien von Testmethoden für Fragebogen unterscheiden: Quantitative Verfahren werden unter Feldbedingungen eingesetzt; dabei werden viele Testpersonen ($N \geq 100$) eingebunden. Im Gegensatz dazu ist die Zahl der Testpersonen bei qualitativen Tests deutlich kleiner. Es werden etwa 15 bis 20 Probanden/Probandinnen ins Pretestlabor eingeladen. Quantitative Tests sind ein guter Weg, um das Ausmaß bestimmter Probleme mit dem Fragebogen zu beurteilen – beispielsweise wie häufig es zu Antwortausfällen oder unplausiblen Antworten kommt. Für eine Überarbeitung von Fragebogen reichen diese Schlussfolgerungen aber oft nicht aus. Demgegenüber geben qualitative Pretests umfassende Einblicke, wie Befragte vorgehen, wenn sie einen Fragebogen beantworten, und zeigen Gründe für Schwierigkeiten auf, die Befragte mit dem Erhebungsinstrument haben.⁶ Auf Basis dieser Erkenntnisse kann der Pretestbereich fundierte Empfehlungen an die entsprechende Fachstatistik geben, wie der Fragebogen überarbeitet werden sollte.

1.3 Drei-Stufen-Ansatz beim qualitativen Pretest

Seit Inbetriebnahme des Pretestlabors arbeitet der Pretestbereich im Statistischen Bundesamt daran, methodische Leitlinien für die Durchführung qualitativer Pretests zu entwickeln, die nach und nach optimiert werden. Im Folgenden wird ein dreistufiger Ansatz vorgestellt, der beschreibt, wie Pretests im Statistischen Bundesamt standardmäßig durchgeführt werden.⁷

⁴ Eurostat: "Handbook on Recommended Practices for Questionnaire Development and Testing in the European Statistical System", Luxemburg 2006.

⁵ Auch andere statistische Ämter haben die Relevanz dieses Themas erkannt, beispielsweise in den Niederlanden. Siehe Snijders, G.: "Cognitive Laboratory Experiences: On Pre-testing Computerised Questionnaires and Data Quality", 2002, Seite 88 f., und Beukenhorst, D.: "The Quality of a Quality Control Instrument: Quality Aspects of Questionnaire Pretesting", Paper auf der European Conference on Quality and Methodology in Official Statistics, Mainz 2004.

⁶ Weitergehende praktische Empfehlungen siehe Blanke, K./Gauckler, B./Sattelberger, S.: „Fragebogen auf dem Prüfstand: Testmethoden und deren Einsatz in der amtlichen Statistik“ in WiSta 8/2008, Seite 641 ff., sowie Radermacher, W./Sattelberger, S.: "Statistical Processes Under Change: Enhancing Data Quality with Pretests" in Fink, A./Lausen, B./Seidel, W./Ultsch, A. (Herausgeber): "Advances in Data Analysis, Data Handling and Business Intelligence", Proceedings der 32. jährlichen Konferenz der Gesellschaft für Klassifikation e.V., Joint Conference with the British Classification Society (BCS) and the Dutch/Flemish Classification Society (VOC), 2010, Seite 67 ff.

⁷ Siehe Radermacher, W./Sattelberger, S. (Fußnote 6). Ein von der Grundidee her ähnlich gelagerter Ansatz mit etwas anderen Stufen (Beobachtung, Follow-up-probing und Debriefing) findet sich bei Hak, T./van der Veer, K./Jansen, H.: "The Three-Step Test-Interview (TSTI): An observation-based method for pretesting self-completion questionnaires" in Survey Research Methods, Jahrgang 2, Nr. 3, 2008, Seite 143 ff.

Stufe 1: Beobachtung während des Ausfüllens⁸

Die Testpersonen werden ins Pretestlabor eingeladen und gebeten, den Fragebogen selbstständig auszufüllen. Um Befragte dabei nicht zu beeinflussen, sind sie allein im Raum, werden aber mit ihrem Einverständnis durch Kameras beobachtet und audiovisuell aufgezeichnet. Auffälligkeiten, die auf Schwierigkeiten mit dem Fragebogen hindeuten, können so gleich notiert werden. Anzeichen für auftretende Unsicherheiten finden sich in kleinen Äußerungen, Mimik und Gestik der Testperson, wenn diese zum Beispiel die Stirn runzelt, sich selbst die Frage laut vorliest oder im Formular mehrmals hin- und herblättert beziehungsweise hin- und herklickt. Beim Test von Online-Fragebogen ermöglicht eine Infrarotkamera, zusätzlich die Blickbewegungen der Testperson live zu verfolgen und aufzuzeichnen (Eye Tracking). Dadurch wird zum Beispiel flüchtiges oder mehrmaliges Lesen am Bildschirm direkt sichtbar. Jedoch sagt die Beobachtung nichts über die Gründe für fehlerhafte oder fehlende Eintragungen aus. Deshalb ist es notwendig, diese Methode mit kognitiven Interviews zu kombinieren.

Stufe 2: Kognitives Interview nach dem Ausfüllen

Nachdem die Testperson den Fragebogen selbstständig ausgefüllt hat, befragt ein geschulter Interviewer oder eine geschulte Interviewerin sie zum Umgang mit dem Erhebungsinstrument.⁹

Zunächst sollen die Testpersonen kurz ihre persönliche Situation bezogen auf das Thema des Fragebogens beschreiben (narrativer Gesprächsbeginn). Im Folgenden wird der Antwortprozess der Testperson durch den Einsatz kognitiver Fragetechniken transparent gemacht. Wie wurden bestimmte Fragen beziehungsweise Antwortkategorien verstanden? Wie kam die Testperson mit dem Layout zurecht? Gleichzeitig wird immer wieder auf den narrativen Gesprächsbeginn Bezug genommen, um zu überprüfen, ob die individuelle Realität in den Fragebogen übertragen werden konnte, beziehungsweise um gegebenenfalls fehlerhafte Eintragungen aufzudecken. Das kognitive Interview wird anhand eines teilstandardisierten Gesprächsleitfadens durchgeführt. Damit soll die Vergleichbarkeit von Ergebnissen für die spätere Analyse über mehrere Testpersonen hinweg gewährleistet werden. Dennoch soll es möglich sein, flexibel auf die jeweilige Befragungssituation einzugehen.

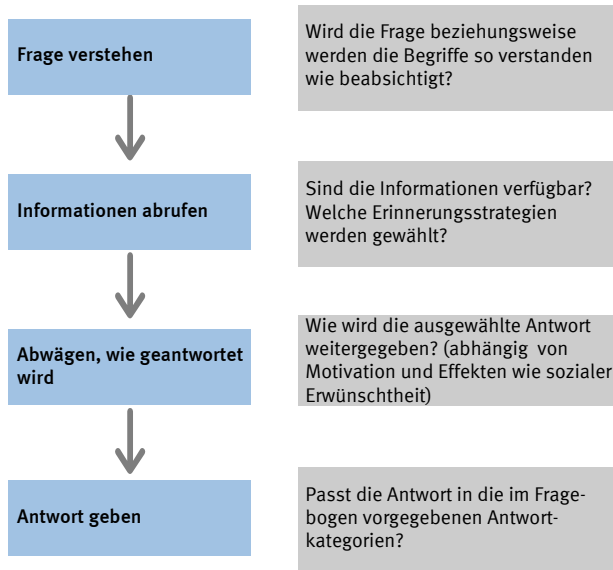
Das kognitive Interview beruht auf Tourangeaus¹⁰ klassischem Modell des Frage-Antwort-Prozesses. Mit einer Frage konfrontiert durchlaufen Befragte im Denkprozess vier miteinander verbundene Schritte (siehe Schaubild 1). Auf allen vier Ebenen können Schwierigkeiten auftreten, die Befragte davon abhalten, korrekte oder vollständige Antworten zu geben.

⁸ Die folgende Beschreibung der ersten Stufe bezieht sich auf den Test von Papier- und Online-Fragebogen für Selbstausfüller. Im Pretestlabor können aber auch Fragebogen, die für Face-to-Face- oder Telefoninterviews vorgesehen sind, getestet werden. In solchen Fällen wird eine Interviewsituation nachgestellt. Die Beobachtung bezieht sich dann auf Interaktionen zwischen interviewter Testperson und Interviewer/-in.

⁹ Auch diese Testsituation wird audio-visuell aufgezeichnet.

¹⁰ Siehe Tourangeau, R./Rips, L.J./Rasinski, K.: "The Psychology of Survey Response", Cambridge 2000, Seite 7 ff.

Schaubild 1 Kognitiver Frage- und Antwortprozess



Quelle: Eigene Abbildung nach Porst, R.: "Fragebogen: Ein Arbeitsbuch.", Wiesbaden 2009, Seite 17ff., und Willis, G. B.: "Cognitive Interviewing: A Tool for Improving Questionnaire Design", Thousand Oaks 2005, Seite 34 ff.

2011 - 01 - 0336

Kognitive Interviewtechniken zielen darauf ab herauszufinden, auf welche Ebene Schwierigkeiten zurückgeführt werden können, sodass der Fragebogen dahingehend nachträglich verbessert werden kann. Beispielsweise bezieht sich die Technik des Paraphrasierens auf den Prozess des Verstehens: Testpersonen werden darum gebeten, eine Frage in eigenen Worten wiederzugeben. Dadurch lässt sich ermitteln, ob die Frage so verstanden und interpretiert wurde, wie beabsichtigt. Beim Test von Online-Fragebogen werden Testpersonen häufig gebeten, ihre Gedanken zu verbalisieren – entweder parallel zum Ausfüllen (Concurrent Think Aloud) oder nach dem Ausfüllen anhand eines Videos der aufgezeichneten Klick- und Blickbewegungen (Retrospective Think Aloud). Durch das laute Denken werden kognitive Prozesse in Worte gefasst.¹¹

Stufe 3: Evaluation des ausgefüllten Fragebogens und gegebenenfalls Auswertung der Eye-Tracking-Daten

Nach dem kognitiven Interview wird der von der Testperson ausgefüllte Fragebogen evaluiert, da dieser wiedergibt, was ursprünglich verstanden und ausgefüllt wurde. Mithilfe der Erkenntnisse aus dem kognitiven Interview wird deutlich, wo es zu Fehleintragungen kam.

Bei Online-Fragebogen kommt eine vertiefte Analyse der Blickdaten, die der Eye Tracker aufgezeichnet hat, hinzu. Zentrale Fragen sind unter anderem: Wo lag die Hauptaufmerksamkeit einzelner oder aller Testpersonen? Welche Ele-

11 Für andere kognitive Techniken (zum Beispiel Probing, Confidence Rating und Sorting) siehe Willis, G. B.: "Cognitive Interviewing: A Tool for Improving Questionnaire Design", Thousand Oaks 2005, Seite 42 ff.

mente des Fragebogens wurden überhaupt nicht beachtet? Wohin fällt der Blick der Testpersonen zuerst?¹²

Der skizzierte dreistufige Ansatz ermöglicht es, durch die Kombination von Beobachtung, kognitivem Interview, Evaluation (und bei Online-Fragebogen von Blickdaten) die Hauptschwierigkeiten mit dem Fragebogen zu identifizieren, Ursachen hierfür aufzudecken und Optimierungsvorschläge für das Erhebungsinstrument zu entwickeln. Welche Maßnahmen getroffen werden, um eine hohe Qualität der Pretest-Ergebnisse zu sichern,¹³ wird im Folgenden dargestellt.

2 Standards zur Qualitätssicherung bei qualitativen Pretests

Ergebnisse, die auf qualitativen Methoden beruhen, werden oft angezweifelt: Sie seien nicht zuverlässig, nicht repräsentativ und nicht signifikant.¹⁴ Im Vergleich zu quantitativen Daten ist die Gefahr tatsächlich höher, subjektiven Interpretationen zu unterliegen, Aussagen einzelner Testpersonen übermäßig zu betonen oder Einzelergebnisse unreflektiert zu verallgemeinern. Um diese Gefahr zu vermeiden, zieht der Pretestbereich im Wesentlichen drei Kriterien nach Miles/Huberman¹⁵ heran, um die hohe Qualität seiner qualitativen Pretests zu sichern:

- (1) Generalisierung: Auswahl von für die Grundgesamtheit typischen Testpersonen und Überprüfung von Aussagen.¹⁶
- (2) Vermeidung von situations- und personenbezogenen Effekten.¹⁷
- (3) Durchführung von Triangulation.

Im Folgenden werden die Kriterien von Miles und Huberman im Einzelnen kurz definiert. Danach wird unter Berücksichtigung der Rahmenbedingungen der amtlichen Statistik skizziert, wie der Pretestbereich diese Kriterien implementiert. Da dieses Vorgehen als "work in progress" betrachtet wird, lassen sich schließlich Verbesserungspotenziale und Zukunftsperspektiven aufzeigen.

12 Für eine detaillierte Beschreibung des Testablaufs für Online-Erhebungsinstrumente siehe Tries, S.: "Usability-Tests von Online-Fragebogen" in Methoden – Verfahren – Entwicklungen, Ausgabe 2/2010, Seite 5 ff. (im Internet unter www.destatis.de, Pfad: Wissenschaftsforum → Methoden und Verfahren).

13 Qualitative Forschung kann nicht mit quantitativer Forschung verglichen werden, wenn es darum geht, gesicherte Aussagen über Daten zu treffen. Die gängigen Qualitätsstandards der quantitativen Forschung greifen nicht und sollen deshalb auch im Rahmen dieses Aufsatzes nicht eingelöst werden. Für eine weiterführende Diskussion siehe Flick, U.: "Qualitative Sozialforschung: Eine Einführung", Reinbek bei Hamburg 2007, Seite 487 ff.

14 Siehe Miles, M. B./Huberman, A. M. (Fußnote 1), hier: Seite 1.

15 Miles/Huberman erwähnen eine Vielzahl verschiedener Kriterien, um Schlussfolgerungen zu ziehen und diese zu verifizieren. Diese würden jedoch über den Rahmen dieses Beitrags weit hinausgehen. Die folgenden Ausführungen beziehen sich auf Miles, M. B./Huberman, A. M. (Fußnote 1), hier: Seite 262 ff., und für weitere Kriterien Seite 267 ff.

16 Hier sprechen Miles/Huberman eigentlich von "Checking for Representativeness". Da dies nicht Repräsentativität im quantitativ-statistischen Sinne (basierend auf Zufallsstichproben) meint, haben die Autorinnen die Benennung des Kriteriums angepasst.

17 Miles/Huberman sprechen hier von "Researcher Effects". Da dieses Kriterium jedoch sowohl Einflüsse vom Interviewer beziehungsweise von der Interviewerin auf die interviewte Person und die Testsituation als auch umgekehrt beschreibt, haben die Autorinnen die Bezeichnung angepasst.

2.1 Generalisierung: Auswahl der Testpersonen und Überprüfung von Aussagen

Im Kontext qualitativer Pretests bedeutet „repräsentativ“, dass – abgeleitet aus den Pretest-Ergebnissen – generelle Annahmen über die Grundgesamtheit (also die Zielgruppe des Fragebogens) und ihre Schwierigkeiten mit dem Formular getroffen werden können. Um Repräsentativität in diesem Sinne zu gewährleisten, sind zwei potenzielle Problemfelder auszuräumen: a) Verzerrungen bei der Rekrutierung von Testpersonen und b) Generalisierung einzelner Aussagen ohne Verifizierung.

- a) Für kognitive oder andere sozialwissenschaftliche Tests werden häufig Studierende befragt. Da Haushaltsbefragungen der amtlichen Statistik sich jedoch häufig an ein breites Spektrum der Bevölkerung richten, ist ein wesentliches Anliegen der Rekrutierungsbemühungen des Pretestlabors im Statistischen Bundesamt, eine Gruppe von 15 bis 20 Testpersonen zusammenzustellen, die sich aus dem gesamten potenziellen Befragtenkreis zusammensetzt. Damit die Grundgesamtheit möglichst gut abgebildet und Probleme mit dem Fragebogen umfassend aufgedeckt werden können, ist es darüber hinaus sinnvoll, dass sich die eingeladenen Testpersonen in einigen entscheidenden Merkmalen (zum Beispiel Alter, Geschlecht, Bildungsgrad oder derzeitiger Erwerbsstatus) unterscheiden.¹⁸

Die Anwerbung von Testpersonen für Haushaltsbefragungen erfolgt beim Statistischen Bundesamt auf verschiedenen Wegen: Es werden Informationsflyer in öffentlichen Institutionen ausgelegt sowie Online- oder Zeitungsanzeigen geschaltet. Testpersonen werden aber auch über persönliche Kontakte (zum Beispiel über andere Probanden/Probandinnen) rekrutiert. Informationen zum persönlichen Hintergrund der bisherigen Testpersonen sind in einer Datenbank gespeichert und werden beständig aktualisiert.¹⁹ Für jeden Pretest muss der Pretestbereich in Absprache mit der Fachstatistik vorab festlegen, welche Testpersonen aus der Datenbank ausgewählt oder ob Testpersonen neu angeworben werden sollen, um später die Pretest-Ergebnisse verallgemeinern zu können.²⁰

Einschränkend ist zu sagen, dass eine leichte Verzerrung in den Pretest-Ergebnissen nicht zu vermeiden ist. Denn bei den Testpersonen handelt es sich ausschließlich um Freiwillige, die entweder einen Einblick in die amtliche Statistik erhalten wollen oder durch die Aufwandsentschädigung von 30 Euro zur Teilnahme motiviert werden. Deren Interesse am Thema der Erhebung ist deshalb in jedem Fall höher einzustufen als bei Befragten der amtlichen Statistik, die in der Regel zur Auskunft verpflichtet sind. Probleme, die im Pretest identifiziert werden (zum

Beispiel falsches Begriffsverständnis aufgrund flüchtigen Lesens), fallen in der Erhebung deshalb vermutlich eher noch deutlicher aus.

- b) Von jeder Testperson stehen nach dem Pretest viele qualitative Daten zur Verfügung. Diese reichen – je nach Testdesign – von codierten Häufigkeiten zum Umgang mit dem Fragebogenlayout über wörtliche Zitate zum Verständnis von Frageinhalten bis hin zu individuellen beziehungsweise aggregierten Blickverläufen. Durch diese Datenmenge besteht die Gefahr, dass bei der Auswertung Informationen reduziert, selektiert und/oder nur vereinfacht wahrgenommen werden.²¹ Hier ist besondere Vorsicht geboten: Schlussfolgerungen, die für eine oder zwei besonders auffallende Testpersonen zutreffen, dürfen nicht unweigerlich als typisch (das heißt repräsentativ) angesehen werden, denn es könnte sich dabei auch um „Ausreißer“ handeln. Generalisierung ohne Verifizierung ist einer der schwerwiegendsten Vorwürfe an die qualitative Sozialforschung. Der Pretestbereich setzt folgende Sicherheitsvorkehrungen ein, um dieses Problem zu lösen:
- › Wenn nur ein oder zwei konkrete Beispiele für Schlussfolgerungen vorliegen oder es bei den meisten Testpersonen keine Ausnahmen und Abweichungen gibt, sind die Pretest-Ergebnisse nochmals zu überprüfen. Im Abschlussbericht, den der Pretestbereich der Fachstatistik vorlegt, ist aufgeführt, wenn eine geringe Anzahl an Testpersonen anders denkt oder sich anders verhält als die meisten anderen Testpersonen; sie werden dem Ergebnis für die Mehrheit gegenübergestellt.
 - › Für die Analyse sind Angaben von Fallzahlen hilfreich, um die große Menge von verfügbaren qualitativen Daten zu strukturieren. Zu diesem Zweck werden die vorliegenden qualitativen Daten mit einer QDA-Software²² dokumentiert und codiert. Bei der Vorbereitung eines Pretests wird hierzu zunächst ein Codesystem erstellt, das auf der Struktur des kognitiven Gesprächsleitfadens basiert. Geschlossene Fragen können entsprechend direkt während der Beobachtung des Interviews codiert werden. Offene Antworten werden teilweise transkribiert, teilweise inhaltlich zusammengefasst und nach Abschluss der Testphase neu entwickelten Codes zugeordnet. Um einen schnellen und umfassenden Überblick über die Pretest-Ergebnisse zu erhalten, werden die Häufigkeiten der codierten Antworten aller Testpersonen herangezogen.

Daneben lassen sich qualitative Daten zur leichteren Handhabung auch in Matrizen ordnen: Matrizen werden verwendet, um Zusammenhänge zwischen mehreren Variablen (und teilweise Subvariablen) darzustellen. Hierbei kann es sich sowohl um zwei- als auch um mehrdimensionale Darstellungen handeln.²³ Zum Bei-

18 Für Unternehmensstatistiken sind andere Merkmale entscheidend, zum Beispiel regionaler Standort, Betriebsgröße, Anzahl der Beschäftigten oder Wirtschaftsbeziehungsweise Produktionszweig.

19 Potenzielle Testpersonen erklären schriftlich ihr Einverständnis, dass ihre persönlichen Informationen für drei Jahre gespeichert werden dürfen; danach werden diese Daten gelöscht, worüber die Testpersonen schriftlich informiert werden.

20 Für Unternehmensbefragungen erfolgt die Anwerbung in der Regel in Zusammenarbeit mit den Statistischen Ämtern der Länder, die oft über direkte Kontakte verfügen.

21 Siehe Miles, M. B./Huberman, A. M. (Fußnote 1), hier: Seite 11.

22 Zur Hilfestellung, die QDA-Software leistet, definiert Weitzman, E. A.: „Software for qualitative data analysis (QDA) allows the analyst to systematically index and organize the data, and then to reliably and flexibly retrieve the data in many different ways“. Weitzman, E. A.: „Advancing the Scientific Basis of Qualitative Research“ in Ragin, C. C./Nagel, J./White, P. (Herausgeber): „Workshop on Scientific Foundations of Qualitative Research“, Seite 145 ff. (im Internet unter www.nsf.gov/pubs/2004/nsf04219/start.htm, abgerufen am 11. Mai 2011).

23 Siehe Miles, M. B./Huberman, A. M. (Fußnote 1), hier: Seite 239.

spiel können die Angaben zu einer bestimmten Frage über alle Testpersonen oder über Subgruppen hinweg aufgeführt und im Kontext des narrativen Gesprächsbegins auf Vollständigkeit sowie Richtigkeit geprüft werden. Durch eine systematische Gegenüberstellung der qualitativen Daten lässt sich bei der Auswertung Selbsttäuschung (zum Beispiel Generalisierung ohne Verifizierung) vermeiden. Die Schlussfolgerungen sind für die Fachstatistik nachzuvollziehen, da sich die Matrizen als Anhang im Abschlussbericht wiederfinden.²⁴

- › Bei der Auswertung muss darauf geachtet werden, dass Fälle unter den Testpersonen, die den ursprünglichen Vorstellungen zum Umgang mit dem Fragebogen widersprechen, nicht beiseitegeschoben werden. Deshalb ist es ideal, festgehaltene Ergebnisse mehrmals und durch verschiedene Personen zu überprüfen. Zum Fragebogen in all seinen Facetten und zu den Antworten der Testpersonen kann man sich nur im Austausch mit anderen ein vollständiges und ausgewogenes Bild machen. Folglich erfordert die Durchführung von Pretests in hohem Maße Teamarbeit: Meinungen zu uneindeutigen Aussagen, die Zuordnungen zu Codes und entsprechende Schlussfolgerungen werden diskutiert, um zu einer gemeinsamen Interpretation zu gelangen.

2.2 Situations- und personenbezogene Effekte

Mit "Researcher Effects" beschreiben Miles/Huberman zwei Szenarien, die bei Feldstudien zu Verzerrungen führen könnten und die auf Pretests (vor allem auf kognitive Interviews) übertragbar sind:²⁵

- a) Das Verhalten der Testperson kann sowohl vom Interviewer beziehungsweise von der Interviewerin als auch durch die Testsituation selbst beeinflusst werden.
 - › Das Pretestlabor ist zunächst ein ungewohnter Ort und allein aufgrund der Namensgebung könnten sich Testpersonen eingeschüchtert fühlen.²⁶ Um den Umgang mit der künstlichen Situation zu erleichtern, sieht das Pretestlabor des Statistischen Bundesamtes wie ein gewöhnlicher Konferenzraum aus; Pflanzen und große Fenster schaffen eine angenehme Atmosphäre. Während des Tests werden Kekse und Getränke angeboten. Kameras und Mikrofone sind kaum wahrnehmbar an der Zimmerdecke angebracht und werden vom Beobachtungsraum aus bedient. Die Testpersonen müssen vorab in schriftlicher Form der audiovisuellen Aufnahme zustimmen. Allerdings wird diese in der Regel nach einer Weile vergessen, da beim Ausfüllen die größte Aufmerksamkeit auf dem Fragebogen liegt. Die Beobachtung wird während des Interviews nicht mehr

explizit erwähnt, um die Testpersonen nicht zu verunsichern.

- › Ein häufiges Phänomen in (qualitativen) Interviews ist die Verzerrung durch soziale Erwünschtheit: Befragte wollen demnach gegenüber dem Interviewer beziehungsweise der Interviewerin ein positives Bild abgeben und berichten deshalb, was vermeintlich von ihnen erwartet wird, auch wenn dies nicht der Realität entspricht.²⁷ Dieses Verhalten kommt in kognitiven Interviews beispielsweise vor, wenn es darum geht, ob Ausfüllhinweise und Erläuterungen beachtet wurden, was Testpersonen in der Regel bejahen, auch wenn Rückfragen oder die Auswertung von Blickbewegungsdaten teilweise eine andere Schlussfolgerung nahelegen. Werden Rückfragen im Interview auffallend häufig und offensichtlich unabhängig vom Frageinhalt mit Ja beantwortet, kann dies auch als Zustimmungstendenz gedeutet werden, welche ebenfalls zu Verzerrungen führt.²⁸ Um diese Effekte zu vermeiden, ist ein erster notwendiger Schritt, den Testpersonen zu Beginn zu erklären, was während des Tests passiert und welche Erwartungen an sie gestellt werden. Die einleitenden Worte (das sogenannte Warming up) könnten daher lauten: „Ich habe diesen Fragebogen nicht entwickelt. Deshalb können wir ganz offen darüber reden. Ich bin an Ihrer ehrlichen Meinung interessiert. Ihre Anmerkungen helfen dabei, diesen Fragebogen zu verbessern. Deshalb sprechen wir auch über falsche oder fehlende Angaben. Machen Sie sich darüber keine Gedanken. Dies liegt nicht an Ihnen, sondern am Fragebogen, der noch überarbeitet werden muss. Der Fragebogen wird getestet, nicht Sie.“²⁹ Im Interview selbst ist es wichtig, die Testpersonen erneut zur Kritik am Fragebogen einzuladen und ihnen Unsicherheiten zu nehmen (genauere Ausführungen zum Interviewerverhalten siehe Abschnitt 2.2 b)].
- › Manche Testpersonen setzen das Statistische Bundesamt automatisch mit anderen staatlichen Einrichtungen gleich und sorgen sich deshalb, dass bestimmte Daten aus dem Fragebogen an den „Staat“ weitergegeben werden (zum Beispiel Antworten zu Einkommensfragen). Die Folge ist, dass die entsprechenden Informationen nicht mitgeteilt beziehungsweise verfälscht werden. Deshalb ist es unerlässlich, zu Beginn des Pretests hervorzuheben, dass einzelne Fälle im Abschlussbericht für die Fachstatistik anonymisiert werden und das Statistische Bundesamt keine Einzeldaten an Dritte weitergeben darf. Auch die im Rahmen von Pretests erhobenen Daten unterliegen in vollem Umfang dem Datenschutz.
- b) Umgekehrt kann das Verhalten von Testpersonen auch bestimmte Reaktionen beim Interviewenden hervorrufen oder die Testsituation insgesamt beeinflussen. Interviewer/-innen unterliegen der Gefahr, sich allzu sehr dem Verhalten der Testperson anzupassen, zum Beispiel

²⁴ Für mehr Informationen zur Grundidee und Ausgestaltung von Matrizen siehe Miles, M. B./Huberman, A. M. (Fußnote 1), hier: Seite 90 ff.

²⁵ Siehe Miles, M. B./Huberman, A. M. (Fußnote 1), hier: Seite 265 f.

²⁶ Bei Unternehmensbefragungen erfolgen die kognitiven Interviews in der Regel vor Ort in den Betrieben (in sogenannten "company-site-visits"), unter anderem weil die Testpersonen häufiger in ihrer Unternehmensdokumentation nachschlagen müssen. Da der Fragebogen vorab ausgefüllt wird, entfällt die Beobachtung. Begleitet wird der Interviewer beziehungsweise die Interviewerin von einer Person aus der Fachstatistik, die bei detaillierten Rückfragen zum Fragebogenschema Auskunft geben kann. Das Gespräch wird mit Einverständnis der Testperson als Audioaufnahme aufgezeichnet.

²⁷ Siehe Häder, M.: „Empirische Sozialforschung: Eine Einführung“, Wiesbaden 2006, Seite 207.

²⁸ Siehe Schnell, R./Hill, P./Esser, E.: „Methoden der empirischen Sozialforschung“, 7. Auflage, München/Wien 2005, Seite 354 f.

²⁹ Eigene Weiterentwicklung basierend auf Willis, G. B. (Fußnote 11), hier: Seite 142 f.

Suggestivfragen zu stellen und ihrem Gegenüber Worte in den Mund zu legen, wenn die Testperson sich offenbar nicht an kognitive Prozesse erinnern oder diese nicht artikulieren kann. Dadurch werden mitunter die in Abschnitt 2.2 a) genannten Verzerrungseffekte der sozialen Erwünschtheit und der Zustimmungstendenz gefördert. Eine gewisse Distanz zur Testperson ist deshalb absolut notwendig.³⁰

- › Interviewer/-innen sollten sich immer das Grundprinzip der professionellen Interviewführung vor Augen halten, nämlich neutral zu sein. Dies bedeutet vor allem, nicht die eigene Meinung zu äußern, keine Hintergrundinformationen zur getesteten Frage weiterzugeben und die Befragten bei der Formulierung ihrer Antworten nicht zu beeinflussen beziehungsweise keine Antwort vorzuzunehmen.³¹ Dabei hilft im kognitiven Interview der teilstandardisierte Leitfaden.
- › Es ist wichtig, dass das Interview nicht zu einer alltäglichen „Plauderei“ wird, sondern dass eine gewisse Professionalität gewahrt bleibt.³² Hierzu ist es beispielsweise notwendig, spezifische Verhaltensweisen von Testpersonen zu kennen (zum Beispiel Verschlossenheit, Arroganz, übermäßige Gesprächigkeit) und angemessen darauf zu reagieren.
- › Um Interviewereffekte zu reduzieren, ist es sinnvoll, je Pretest mindestens zwei Interviewer/-innen einzusetzen, die erfahren und gut ausgebildet sind. Im Idealfall geben sich diese wechselseitig Rückmeldung dazu, wie sie ihr Interviewerverhalten verbessern können (hierzu werden gegebenenfalls auch die vorliegenden Audio- oder Video-Aufzeichnungen genutzt).

2.3 Triangulation

Um Pretest-Ergebnisse zu bestätigen (oder gegebenenfalls zu widerlegen), werden sie im Rahmen der Triangulation aus verschiedenen Perspektiven nachvollzogen. Diese Art von Verifizierung (Bestätigung) beziehungsweise Falsifizierung (Widerlegung) kann auf unterschiedliche Weise erreicht werden.³³ An einigen Stellen dieses Beitrags wurden bereits verschiedene Ansätze der Triangulation beschrieben, ohne explizit darauf aufmerksam zu machen. Verglichen werden hierbei qualitative Daten, die a) aus verschiedenen Quellen stammen, b) unter Einsatz verschiedener Methoden erhoben und c) bei der Auswertung von mehreren Personen bewertet wurden. Der folgende Abschnitt beschreibt, wie Triangulation bei qualitativen Pretests im Statistischen Bundesamt realisiert wird:³⁴

- a) *Datentriangulation* bedeutet, verschiedene Quellen einzubeziehen (zum Beispiel unterschiedliche Testpersonen

oder andere räumliche Rahmenbedingungen/Voraussetzungen). Um die Zielgruppe des Fragebogens möglichst genau abzubilden, werden zu Pretests unterschiedliche Typen potenzieller Befragter eingeladen. Für den Zensus 2011 beantworteten zum Beispiel sowohl ältere Personen als auch Studierende teilweise einen Fragebogen für Bewohner/-innen von Wohnheimen. Bei Pretests für landwirtschaftliche Erhebungen vergleicht der Pretestbereich Betriebe, die in unterschiedlichen Bundesländern angesiedelt sind, weil regionale Unterschiede (zum Beispiel bei der Verwendung von Fachbegriffen) auftreten.

- b) *Methodentriangulation* ist die Kombination mehrerer Methoden, wie sie im Drei-Stufen-Ansatz bei qualitativen Pretests (siehe Abschnitt 1.3) thematisiert wurde. Die angewandten Methoden haben jeweils ihre Stärken und Schwächen. Deshalb zielt der Ansatz darauf ab, die Nachteile einer Methode durch die Vorteile einer anderen Methode auszugleichen. Beispielsweise können über die Beobachtung und die Evaluation des ausgefüllten Fragebogens keine Gründe für Probleme mit dem Fragebogen identifiziert werden, wohingegen das kognitive Interview diesen Erkenntnisgewinn bringt. Umgekehrt können die Ergebnisse aus dem kognitiven Interview durch situations- und personenbezogene Effekte verzerrt oder in ihrer Aussagekraft begrenzt sein, wenn Testpersonen kognitive Prozesse unzureichend wiedergeben, sei es aus Erinnerungs- oder Verbalisierungsschwierigkeiten. Notwendig ist daher, die Erkenntnisse aus dem Interview durch die beiden anderen, weniger reaktiven Methoden (Beobachtung und Evaluation des Fragebogens) zu bestätigen.
- c) *Forschertriangulation* bezieht sich nicht auf einfache Arbeitsteilung, sondern auf einen systematischen Vergleich von Eindrücken unterschiedlicher Personen, die an der Durchführung und Auswertung beteiligt sind. Im Pretestbereich erfolgt die entsprechende Umsetzung dadurch, dass mehrere Personen mit Interviews, Codierung und Beobachtung betraut sind. Wie bereits betont wurde, implizieren qualitative Pretests Teamarbeit und wechselseitiges Feedback, was unerlässlich ist, um subjektive Sichtweisen auf den Fragebogen zu vermeiden. Während des Pretests und der Datenanalyse findet deshalb ein kontinuierlicher Meinungs- und Erfahrungsaustausch (auch mit der Fachstatistik) statt.

Insgesamt besteht das Ziel der Triangulation darin, Informationen aus unterschiedlichen Perspektiven zu sammeln und daraus ein Gesamtbild zu erstellen, das auf eine bestimmte Schlussfolgerung hindeutet und/oder andere Schlussfolgerungen ausschließt.³⁵ Idealerweise ergänzen sich die unterschiedlichen Sichtweisen oder bestätigen sich gegenseitig und liefern damit einen facettenreichen Blick auf den Fragebogen. Jedoch bringen unterschiedliche Perspektiven nicht immer ähnliche Ergebnisse hervor. Wird im kognitiven Interview zum Beispiel gefragt, wie mit den Ausfüllhinweisen umgegangen wurde, bejaht die Mehrheit der Testpersonen, diese gelesen zu haben [oftmals bedingt durch soziale Erwünschtheit, siehe Abschnitt 2.2 a)]. Blickverlaufsdaten aus dem Eye Tracking können den Wahrheitsgehalt dieser

30 Solche Effekte treten auch in der Interpretation der Daten auf. Zum Beispiel beeinflussen persönliche Einstellungen der Auswertenden die Schlussfolgerungen während der Analyse. Erwähnt wird diese Art von Verzerrung unter Abschnitt 2.1 (Generalisierung ohne Verifizierung).

31 Prüfer, P./Stiegler, A.: „Die Durchführung standardisierter Interviews: Ein Leitfaden.“, Mannheim 2002, Seite 3.

32 Siehe Prüfer, P./Stiegler, A. (Fußnote 31), hier: Seite 4.

33 Zur klassischen Ausdifferenzierung siehe Denzin, N. K.: „The Research Act“, Chicago 1970.

34 Basierend auf Flick, U.: „Triangulation: Eine Einführung“, 2. Auflage, Wiesbaden 2008, Seite 13 ff.

35 Siehe Miles, M. B./Huberman, A. M. (Fußnote 1), hier: Seite 266 f.

Aussage prüfen und zeigen in vielen Fällen, dass die Ausföhlhinweise komplett überlesen wurden. Wenn bei der Analyse aufkommende Widersprüche nicht geklärt werden können, nimmt der Pretestbereich diese in seinem Abschlussbericht an die Fachstatistik als offene Punkte auf, die in weiteren Pretests geklärt werden sollten.

- › Ein intensiver Erfahrungsaustausch mit anderen statistischen Ämtern und Institutionen, die ebenfalls Pretests durchführen, trägt dazu bei, laufende Prozesse zu überdenken und zu verbessern. [\[4\]](#)

3 Schlussfolgerungen

Der vorliegende Beitrag beschreibt, wie das Statistische Bundesamt mithilfe eines dreistufigen Ansatzes Fragebogen für Selbstausfüller testet. Die hohe Qualität der während des Pretests gesammelten qualitativen Daten wird gewährleistet, indem Generalisierung ohne Verifizierung ausgeschlossen wird, für die Grundgesamtheit typische Testpersonen ausgewählt werden, situations- und personenbezogene Effekte vermieden werden und die Triangulation umfassend durchgeführt wird. Zusammenfassen lassen sich die wichtigsten Sicherheitsvorkehrungen für aussagekräftige Pretest-Ergebnisse, die letztlich die Grundlage für eine fundierte Fragebogenoptimierung sind, wie folgt:

- › Qualitative Pretests implizieren Teamarbeit. Die Verteilung verschiedener Aufgaben (zum Beispiel beobachten, codieren, kognitive Interviews führen, den Abschlussbericht schreiben) auf mehrere Personen im Pretestbereich ist notwendig, um zu einer gemeinsamen, weniger subjektiven Interpretation zu kommen. Jedoch funktioniert Teamarbeit nur dann gut, wenn die Beteiligten sich immer wieder austauschen und gegenseitig Rückmeldung geben.
- › Dadurch, dass während der Datensammlung und -analyse unterschiedliche Sichtweisen auf den Fragebogen eingenommen und verglichen werden, werden die Pretest-Ergebnisse auf effiziente Art und Weise immer wieder gründlich überprüft und dabei im Idealfall bestätigt (Triangulation).
- › Der Einsatz von QDA-Software und Matrizen hilft dabei, die große Menge an qualitativen Daten zu strukturieren. Die transparente Dokumentation der Analyseschritte ist wichtig, um Selbsttäuschung zu vermeiden und die Pretest-Ergebnisse für die Fachstatistik nachvollziehbar zu machen.

Auch künftig wird die hohe Qualität von Pretest-Ergebnissen ein zentrales Anliegen des Statistischen Bundesamtes sein. Folgende Aspekte werden in künftige Entwicklungen einbezogen:

- › Standards zur Evaluierung der Qualität qualitativer Pretests müssen definiert und weiterentwickelt werden, vor allem was Online-Fragebogen und Unternehmensstatistiken betrifft.
- › Neue Rekrutierungsmöglichkeiten sind zu erproben, um potenzielle Befragte anzusprechen, die bisher unterrepräsentiert sind (zum Beispiel Personen mit Migrationshintergrund und Nicht-Muttersprachler/-innen).
- › Detaillierte Leitlinien für kognitive Interviews helfen dabei, auf schwierige Testsituationen angemessen zu reagieren.

Auszug aus Wirtschaft und Statistik

Herausgeber

Statistisches Bundesamt, Wiesbaden

www.destatis.de

Schriftleitung

Roderich Egeler, Präsident des Statistischen Bundesamtes

Brigitte Reimann (verantwortlich für den Inhalt)

Telefon: + 49 (0) 6 11 / 75 20 86

Ihr Kontakt zu uns

www.destatis.de/kontakt

Statistischer Informationsservice

Telefon: + 49 (0) 6 11 / 75 24 05

Telefax: + 49 (0) 6 11 / 75 33 30

Abkürzungen

WiSta	=	Wirtschaft und Statistik
MD	=	Monatsdurchschnitt
VjD	=	Vierteljahresdurchschnitt
HjD	=	Halbjahresdurchschnitt
JD	=	Jahresdurchschnitt
D	=	Durchschnitt (bei nicht addierfähigen Größen)
Vj	=	Vierteljahr
Hj	=	Halbjahr
a. n. g.	=	anderweitig nicht genannt
o. a. S.	=	ohne ausgeprägten Schwerpunkt
St	=	Stück
Mill.	=	Million
Mrd.	=	Milliarde

Zeichenerklärung

p	=	vorläufige Zahl
r	=	berichtigte Zahl
s	=	geschätzte Zahl
–	=	nichts vorhanden
0	=	weniger als die Hälfte von 1 in der letzten besetzten Stelle, jedoch mehr als nichts
.	=	Zahlenwert unbekannt oder geheim zu halten
...	=	Angabe fällt später an
X	=	Tabellenfach gesperrt, weil Aussage nicht sinnvoll
I oder —	=	grundsätzliche Änderung innerhalb einer Reihe, die den zeitlichen Vergleich beeinträchtigt
/	=	keine Angaben, da Zahlenwert nicht sicher genug
()	=	Aussagewert eingeschränkt, da der Zahlenwert statistisch relativ unsicher ist

Abweichungen in den Summen ergeben sich durch Runden der Zahlen.