

Kognitive Interviews

Peter Prüfer & Margrit Rexroth

Zentrum für Umfragen, Methoden und Analysen, Mannheim

Zusammenfassung

In der Umfrageforschung sind kognitive Interviews ein effektives Werkzeug, um potentielle Probleme bei Survey-Fragen zu identifizieren. In diesem Beitrag werden die wichtigsten kognitiven Techniken vorgestellt und Empfehlungen für die Durchführung kognitiver Interviews gegeben.

**ZUMA How-to-Reihe, Nr. 15
2005**

Inhaltsverzeichnis

| | Seite |
|--|-------|
| 1. Was sind kognitive Interviews? | 3 |
| 2. Was soll getestet werden? | 4 |
| 3. Die wichtigsten kognitiven Techniken | 5 |
| 3.1 Nachfragen | 5 |
| <i>Nachfragen zum Verständnis</i> | 6 |
| <i>Nachfragen zur Wahl der Antwortkategorie/des Skalenwertes</i> | 7 |
| <i>Nachfragen zum Besitz relevanter Informationen</i> | 8 |
| <i>Nachfragen zur Informationsbeschaffung/-gewinnung</i> | 9 |
| <i>Unspezifische Nachfragen</i> | 9 |
| 3.2 Technik zur Bewertung der Verlässlichkeit einer Antwort..... | 10 |
| 3.3 Paraphrasieren | 12 |
| 3.4 Sortiertechnik | 13 |
| 3.5 Technik des lauten Denkens | 14 |
| 4. Empfehlungen zur Durchführung kognitiver Interviews | 16 |
| 5. Wie wertet man kognitive Interviews aus? | 18 |
| 6. Finanzieller und zeitlicher Aufwand | 18 |
| Literatur | 19 |
| Beispiel für ein Auflistungsschema zur Auswertung einer Frage über alle Interviews | 21 |

1. Was sind kognitive Interviews?

Kognitive Interviews führt man in der Entwicklungsphase eines Fragebogens durch, um Einblick in die kognitiven Prozesse zu bekommen, die beim Beantworten von Fragen ablaufen. Speziell ist dabei von Interesse, wie Befragte

- Fragen oder Begriffe interpretieren und verstehen,
- Informationen und Ereignisse aus dem Gedächtnis abrufen,
- Entscheidungen darüber treffen, wie sie antworten,
- ihre „intern“ ermittelte Antwort formalen Antwortkategorien zuordnen.

Das eigentliche Ziel dabei besteht darin, durch diese Informationen Hinweise auf unterschiedlichste Frage-Probleme zu erhalten.

Wie kann man nun Informationen über die genannten Aspekte erhalten?

Mit einem einfachen Beobachtungs-Pretest¹, bei dem der Testfragebogen unter den Bedingungen der Hauptstudie² erhoben wird und dabei lediglich die von den Testpersonen spontan geäußerten Bemerkungen registriert werden, ist dies nur in geringem Umfang und zudem nur sehr unsystematisch möglich, und zwar vor allem deshalb, weil Befragte Probleme bei der Beantwortung einer Frage von sich aus nur selten äußern. Dies ist häufig darauf zurückzuführen, dass Befragte in „Problem-Fällen“ eine Frage für sich selbst so modifizieren, dass sie sie leichter beantworten können.³

Zum Beispiel beantwortet ein Befragter die Frage „Wie oft haben Sie in den letzten vier Wochen in einem Supermarkt eingekauft?“ nur für die beiden letzten Wochen, an die er sich besonders gut erinnern kann, die ersten zwei Wochen berücksichtigt er bei seiner Antwort nicht. Er verändert die Frage so, dass sie für ihn leichter zu beantworten ist.

Probleme dieser Art können also mit einer Vorgehensweise wie beim Beobachtungs-Pretest nicht zuverlässig aufgedeckt werden, sondern nur mit kognitiven Interviews, bei denen gezielt Nachfragen bzw. andere „aktive“ Techniken eingesetzt werden.

Kognitive Interviews zeichnen sich durch folgende Merkmale aus (Praktische Hinweise zur Durchführung kognitiver Interviews werden in Kapitel 4, Seite 16 gegeben):

- Kognitive Interviews werden durchgeführt, um Informationen über Frage-Probleme zu erhalten.
- Zur Identifikation von Frage-Problemen werden in kognitiven Interviews sog. kognitive Techniken eingesetzt, z.B. Nachfragen oder die Technik des lauten Denkens (siehe Kapitel 3, Seite 5).
- Die Testprozedur läuft in einem kognitiven Interview normalerweise in zwei Schritten ab: Im ersten Schritt wird der Testperson die zu testende Frage gestellt. Nach deren Beantwortung stellt der Testleiter dann Nachfragen oder setzt andere kognitive Techniken ein. Eine Ausnahme bildet die Technik des lauten Denkens, die bereits vor der Beantwortung der zu testenden Frage einsetzt.

¹ Siehe z.B. Prüfer, P./Rexroth, M., 1996

² Der Fragebogen wird in der Form erhoben, wie er auch in der Hauptstudie zum Einsatz kommt. Die Testpersonen sind nicht über den Testcharakter des Interviews informiert.

³ Siehe dazu Belson, W.A., 1981

- Kognitive Interviews werden normalerweise persönlich-mündlich durchgeführt. Der Interviewer wird hier Testleiter⁴ genannt, die Interviewten werden - im Unterschied zu „echten“ Befragten - als Testpersonen bezeichnet.
- Mit kognitiven Interviews werden in erster Linie einzelne Fragen getestet; sie eignen sich nicht dazu, die Funktionsfähigkeit eines fertig gestellten Fragebogens zu überprüfen, z.B. auf fehlerfreien Ablauf einer Filterführung oder bezüglich seiner Handhabbarkeit durch Interviewer.

2. Was soll getestet werden?

Wenn Fragen kognitiv überprüft werden sollen, gilt es zunächst zu klären, welche potentiellen Probleme bei jeder einzelnen Frage auftreten könnten.

Dabei sollte man unterscheiden zwischen „eher technischen“ Problemen, wie z.B. fehlende Anweisungen oder eine falsche Filterführung. Solche Probleme können leicht ohne empirische Überprüfung „am Schreibtisch“ gelöst werden und sind nicht Gegenstand der Überprüfung mittels kognitiver Interviews.

Probleme, die mit kognitiven Interviews aufgedeckt werden können, beziehen sich auf den Prozess, der bei der Beantwortung einer Frage modellhaft in vier Stufen abläuft⁵. Die Stufen des Antwortmodells können bei der Ermittlung möglicher Probleme dabei als Hilfe dienen:

1. Verständnis der Frage/von Begriffen

Frageinhalt:

Was meint der Befragte, was mit der Frage gefragt wird?

Bedeutung von Begriffen:

Wie versteht der Befragte bestimmte Begriffe bzw. Sätze?

2. Abruf relevanter Informationen bzw. Ereignisse aus dem Gedächtnis

Abrufbarkeit von Informationen:

Welche Arten von Informationen benötigt der Befragte?

Erinnerungsstrategien:

Welche Erinnerungsstrategie wird eingesetzt?

Versucht der Befragte z.B. zu schätzen oder sich ganz detailliert an Dinge zu erinnern?

Gibt es eventuelle Probleme mit der Erinnerungsfähigkeit bzw. dem Vorhandensein von notwendigen Informationen?

3. Entscheidungen darüber treffen, wie geantwortet wird

Motivation:

Gibt sich der Befragte Mühe bei der Beantwortung der Frage?

Soziale Erwünschtheit:

Sagt der Befragte die Wahrheit?

Versucht er, sich besser darzustellen?

Ist der Befragte willens, die Frage ehrlich zu beantworten?

4. Zuordnung der „intern“ ermittelten Antwort zu formalen Antwortkategorien

Kann der Befragte seine „intern“ ermittelte Antwort den formalen Antwortkategorien zuordnen?

Zum Erkennen möglicher Probleme liefert dieses hier vorgestellte 4-Stufen-Modell nur grobe Anhaltspunkte. Eine konkrete Hilfestellung gibt ein System zur Bewertung der Fragenqualität (FBS) von Faulbaum, Prüfer und Rexroth (2009). In diesem System werden sehr detailliert und anhand vieler Beispiele aus der Umfragepraxis Hinweise auf mögliche Probleme beim Verständnis und bei der Beantwortung von Fragen gegeben.

⁴ Die männliche Form von „Interviewer“, „Testleiter“ und „Befragter“ wird in diesem Bericht lediglich der Einfachheit halber verwendet.

⁵ Siehe dazu ein Modell des Frage-Antwort-Prozesses von Tourangeau, 1984

3. Die wichtigsten kognitiven Techniken

In diesem Kapitel werden fünf kognitive Techniken vorgestellt, mit denen potentielle Probleme identifiziert werden können. Es sind Techniken, die sich bei ZUMA in der praktischen Arbeit mit kognitiven Interviews bewährt haben und/oder die in der einschlägigen Literatur häufig genannt werden:

- **Nachfragetechniken (Probing)**
- **Bewertung der Verlässlichkeit der Antwort (Confidence Rating)**
- **Paraphrasieren (Paraphrasing)**
- **Sortiertechniken (Card Sorting)**
- **Technik des lauten Denkens (Think Aloud)**

Im Folgenden wird jede Technik kurz beschrieben und anhand von Beispielen demonstriert, wie sie eingesetzt wird und welche Informationen man mit ihrem Einsatz erhalten kann. Die Beispiele stammen aus kognitiven Interviews, die alle im kognitiven Labor von ZUMA stattfanden.

3.1 Nachfragen (Probing)

Nachfragetechniken nehmen unter den kognitiven Techniken eine zentrale Rolle ein. Mit Hilfe dieser Techniken wird mittels einer oder mehrerer Zusatzfragen das Verständnis der ganzen Frage oder von Elementen der Frage gezielt hinterfragt. Zu diesen Elementen zählen z.B. einzelne Worte oder Begriffe, vorgegebene Antwortkategorien oder das Verständnis von Ausfüllanweisungen bei einem schriftlichen Fragebogen.

Das Nachfragen (Probing) ist eine altbewährte Interview-Technik, die bereits von Schuman (1966) und Belson (1981) eingesetzt wurde, um zu überprüfen, ob Befragte eine Frage im Sinne des Fragenkonstruktors verstanden bzw. beantwortet haben.

Es gibt verschiedene Varianten von Nachfragetechniken:

- **Nachfragen zum Verständnis (Comprehension Probing)**

Bei Nachfragen zum Verständnis sollen Befragte beschreiben, wie sie bestimmte Aspekte einer Frage oder eines Begriffs verstehen. Dabei stellt der Testleiter der Testperson die Nachfragen in der Regel nach der Beantwortung der eigentlichen Frage.

Beispiel für eine Nachfrage zum Verständnis

Fragetext:

„Inwieweit stimmen Sie der folgenden Aussage zu: Meine Gesundheit ist in der Hauptsache eine Frage von guter Anlage und Glück.“

Skala auf Liste: stimme voll und ganz zu = 1/stimme eher zu = 2/teils teils = 3/lehne eher ab = 4/lehne voll und ganz ab = 5

Nachfrage zum Verständnis des Begriffs „Glück“ (zu stellen nach Beantwortung der Frage):

„Wie haben Sie das Wort „Glück“ verstanden?“

Antwortbeispiele:

Testperson 1:

„Dass man zufrieden ist.“

Testperson 2:

„Wie es mir geht; Freude ging mir dabei durch den Kopf.“

Testperson 3:

„Als Schicksal.“

Testperson 4:

„Im Sinne von: Was man von der Vererbung mitbekommen hat, erbliche Vorbelastung, also eher Zufall.“

Befund:

Die Frage zielt darauf ab, den Begriff „Glück“ im Sinne von „Zufall“ zu interpretieren. Auf Grund der Nachfrage wird deutlich, dass der Begriff aber auch im Sinne von „Glücklichsein, Zufriedenheit“ verstanden wird. Die Ergebnisse machen deutlich, dass die Testpersonen bei der Beantwortung der Frage von einem unterschiedlichen Begriffsverständnis ausgegangen sind.

Empfehlung:

Um ein unterschiedliches Begriffsverständnis zu vermeiden, muss der Begriff „Glück“ entweder präzisiert oder ersetzt werden (z.B. durch den Begriff „Zufall“).

- **Nachfragen zur Wahl der Antwortkategorie/des Skalenwertes (Category Selection Probing)**

Die Technik wird eingesetzt, nachdem Testpersonen die Frage mittels Zuordnung zu einer Antwortkategorie oder durch Vergabe eines Skalenwertes beantwortet haben. Dabei wird gefragt, warum Befragte eine bestimmte Antwortkategorie oder einen bestimmten Skalenwert gewählt haben.

Die Technik eignet sich zur Überprüfung, ob Befragte eine Frage möglicherweise nicht verstanden haben, obwohl sie diese ohne Hinweise auf irgendwelche Probleme im Sinne der Fragentechnik und des Fragenziels, also formal adäquat beantwortet haben. Solche auf Anhieb nicht erkennbaren Unsicherheiten in den Befragtenantworten finden sich häufig bei Survey-Umfragen insbesondere bei Fragen, die durch die Wahl eines Skalenwertes beantwortet werden sollen.

Beispiel für eine Nachfrage zur Wahl eines Skalenwertes

Fragetext:

„Inwieweit stimmen Sie folgender Aussage zu:

Der Arbeitgeberanteil zur Krankenversicherung sollte abgeschafft werden; zum Ausgleich sollten die erwerbstätigen Krankenversicherten eine Lohnerhöhung in der Höhe des derzeitigen Arbeitgeberanteils bekommen.“

Skala auf Liste: 1 = stimme überhaupt nicht zu/ 6 = stimme voll und ganz zu
(Vermutetes Problem: Fragetext zu kompliziert)

Antwortbeispiel 1:

Testperson nannte Skalenwert „4“

Nachfrage:

„Warum haben Sie den Skalenwert “4” gewählt?“

Antwort auf die Nachfrage:

“Warum sollten die Krankenversicherten Geld bekommen? Finde ich nicht okay. Ich habe den Wert 4 gewählt, weil der ungefähr mittendrin ist.“

Antwortbeispiel 2:

Testperson nannte Skalenwert „2“

Nachfrage:

„Warum haben Sie den Skalenwert „2“ gewählt?“

Antwort auf die Nachfrage:

„Ich habe die Frage nicht ganz verstanden. Ich glaube, ich habe Arbeitgeber mit Arbeitnehmer verwechselt.“

Befund:

In Beispiel 1 stimmt die Testperson mit Skalenwert „4“ der Aussage leicht zu, in ihrer Begründung für den Skalenwert reagiert sie jedoch eher ablehnend. In Beispiel 2 äußert die Testperson explizit, dass sie die Frage nicht verstanden hat, obwohl sie mit Wert „2“ geantwortet hatte. Geht man davon aus, dass die Nachfrage nicht gestellt worden wäre, so wären die vorhandenen Verständnisprobleme in beiden Fällen nicht sichtbar geworden, weil die Testpersonen durch die Nennung des Skalenwertes ein formal korrektes Antwortverhalten gezeigt haben, ohne auf ihre Verständnisschwierigkeiten mit dem Fragetext hinzuweisen. Erst die Nachfrage zeigte, dass die genannten Skalenwerte nicht valide sind.

Empfehlung:

Unser Vorschlag ist, den Satzbau zu vereinfachen und die Substantive zu reduzieren. Eine mögliche Variante könnte sein: *„Der Arbeitgeber sollte keinen finanziellen Beitrag mehr zur Krankenversicherung leisten, sondern diese Summe als Lohnerhöhung direkt an die Erwerbstätigen bezahlen.“*

Letztendlich muss der Forscher entscheiden, ob mit diesen Veränderungen die inhaltliche Zielsetzung der Frage noch gewährleistet ist. Wir empfehlen, solche größeren Veränderungen einem weiteren Test zu unterziehen.

- **Nachfragen zum Besitz relevanter Informationen**

Die Technik hat zum Ziel, die gegebene Antwort dahingehend zu hinterfragen, ob die Testperson über ausreichendes Wissen/Informationen bezüglich des abgefragten Inhaltes verfügt. Auch hier wird – wie bei den zuvor dargestellten Probingtechniken – nach der Beantwortung der Frage eine Nachfrage gestellt.

Beispiel für eine Nachfrage zum Besitz relevanter Informationen

Aussage, die bewertet werden soll:

„Es ist ungerecht, dass Israel den Palästinensern Land wegnimmt.“

Skala auf Liste: 1 = stimme überhaupt nicht zu/ 7 = stimme voll und ganz zu

(Vermutetes Problem: Das nötige Hintergrundwissen fehlt, um die Aussage verlässlich zu bewerten.)

Antwort:

Testperson nannte Skalenwert „4“

Nachfrage nach Beantwortung der Frage:

„Wie sind Sie vorgegangen als Sie die Frage beantwortet haben?“

Antwort:

„Ach Gott, ach Gott. Ist das so? Ich weiß nicht, nur so vom Gefühl her.“

Befund:

Die Reaktion auf die Nachfrage zeigt, dass der Skalenwert „4“ eher als eine „weiß nicht“-Antwort zu werten ist. Ohne Nachfrage hätte es allerdings keinen Hinweis auf die Unsicherheit der Testperson hinsichtlich des hier angesprochenen politischen Sachverhalts gegeben.

Empfehlung:

Diese Frage kann eigentlich nur an Personen gestellt werden, die über das entsprechende Faktenwissen verfügen, um ein fundiertes Urteil abgeben zu können. Um zu vermeiden, dass sich Befragte ohne das nötige Faktenwissen einfach einem Skalenwert zuordnen, muss eine Zusatzkategorie wie z.B. „Kann ich nicht beurteilen“ explizit vorgegeben werden.

- **Nachfragen zur Informationsbeschaffung/-gewinnung (Information Retrieval Probing)**

Die Technik wird bei retrospektiven Fragen eingesetzt und soll verdeutlichen, wie Befragte vorgehen, um sich an Informationen zu erinnern. Durch die Art des Vorgehens bei dem Prozess der Rückerinnerung erhofft sich der Forscher Hinweise darauf, inwieweit die gegebene Antwort verlässlich oder geschätzt ist und welche Probleme bei der Beantwortung der Frage auftraten.

Beispiel für eine Nachfrage zur Informationsbeschaffung

Fragetext:

„Wie oft waren Sie im letzten Jahr beim Arzt?“

(Vermutete Probleme: Der Zeitraum ist nicht eindeutig formuliert. Die Testpersonen können sich nicht genau erinnern.)

Nachfrage nach Beantwortung der Frage:

„Wie haben Sie sich daran erinnert, wie sind Sie vorgegangen?“

Antwortbeispiel 1:

„Ich habe im Januar angefangen und dann Monat für Monat überlegt, wie oft ich beim Arzt war, aber an manche Monate kann ich mich nicht mehr richtig erinnern.“

Antwortbeispiel 2:

„Für die letzten 3 Monate kann ich es genau sagen, davor kann ich mich nicht mehr so genau erinnern.“

Antwortbeispiel 3:

„Soll ich die letzten 12 Monate nehmen oder das Jahr 2004?“

Befund:

Die Antwortreaktionen zeigen, dass...

1. der Zeitrahmen für eine verlässliche Bestimmung der Arztbesuche zu groß ist,
2. die Formulierung „im letzten Jahr“ unterschiedlich definiert werden kann.

Empfehlung:

Wir empfehlen, die letzten drei Monate abzufragen, da nach unserer Erfahrung dieser Zeitraum noch gut überblickt werden kann: „Wie oft waren Sie in den letzten drei Monaten beim Arzt?“

- **Unspezifische Nachfragen (General Probing)**

Bei unspezifischen Nachfragen sollen Testpersonen angeben, ob sie bei der Beantwortung einer Frage Probleme hatten. Wenn ja, wird nachgefragt, um welche Probleme es sich im Einzelnen handelt.

Nachfragen allgemeiner Art sind z.B.:

- „Gibt es etwas, was Sie bei der Frage nicht verstanden haben?“
- „Gibt es noch andere Probleme, die wir bis jetzt bei der Frage noch nicht besprochen haben?“

Unspezifische Nachfragen zählen unserer Ansicht nach nicht zu den effektivsten Nachfragetechniken, denn sie müssen nicht in jedem Falle ergiebig arbeiten. So kann die Nachfrage: „Gibt es bei dieser Frage etwas, das Sie nicht verstanden haben?“ nur dann von einer Testperson bejaht werden, wenn sie sich über ihr Nichtverständnis tatsächlich im klaren ist. In vielen Fällen sind sich Befragte eines falschen Fragenverständnisses aber nicht bewusst. Sie können eine mögliche Diskrepanz zwischen ihrem eigenen Verständnis und dem des Forschers

nicht wahrnehmen, da sie das vom Forscher intendierte nicht kennen. Aus diesem Grund decken unspezifisch formulierte Nachfragen weniger Schwierigkeiten bei Fragen auf als spezifische Nachfragen wie z.B. ein gezieltes Nachfragen zum Verständnis eines bestimmten Begriffs. Trotzdem kann man auf unspezifische Nachfragen nicht verzichten. Bei einer Frage sollte jede spezifische Nachfrageprozedur vom Testleiter immer mit einer unspezifischen Nachfrage abgeschlossen werden, um der Testperson die Möglichkeit zu geben, noch andere Probleme oder Unverständlichkeiten als die vom Testleiter bisher angesprochenen formulieren zu können.

Unsere Empfehlung für den Einsatz von Nachfragetechniken:

Mit Ausnahme der unspezifischen Nachfragen sind die Nachfragetechniken – insbesondere das Nachfragen zum Verständnis der gesamten Frage als auch zum Verständnis einzelner Begriffe – hinsichtlich des Informationsgewinns sehr effektive Techniken, weil die Testperson angehalten wird, zu einem ganz bestimmten Sachverhalt Stellung zu nehmen. Generell gilt:

Je präziser die Testperson auf den zu klärenden Inhalt oder Begriff angesprochen wird, desto größer ist die Chance, spezifische Verständnisprobleme aufzudecken.

Die Problematik liegt oftmals in der Formulierung einer spezifischen Nachfrage zum inhaltlichen Verständnis, wenn es nicht gerade um die Überprüfung eines einzelnen Begriffs geht. Zum einen kann man eine effektiv arbeitende Nachfrage nur in genauer Kenntnis der inhaltlichen Zielsetzung der zu überprüfenden Frage konstruieren, zum anderen muss eine Nachfrage selbst valide sein. „Gute“ Nachfragen im Vorfeld zu entwickeln, ist also nicht immer einfach, und es empfiehlt sich, die im Vorfeld konstruierten Nachfragen selbst einem kurzen Test zu unterziehen.

3.2 Technik zur Bewertung der Verlässlichkeit einer Antwort (Confidence Rating)

Die Technik setzt man zur Überprüfung von Faktfragen ein, insbesondere von Rückerinnerungsfragen. So kann man z.B. eine Antwort auf die Frage:
„Seit wie vielen Jahren leben Sie in dieser Stadt?“ mit einer Nachfrage überprüfen.

Dabei gibt es zwei Varianten der Vorgehensweise:

Variante 1:

Die Verlässlichkeitsfrage „*Und wie genau ist Ihre Antwort?*“ wird als offene Frage gestellt, d.h. die Testperson beantwortet die Frage mit eigenen Worten.

Variante 2:

Bei der Verlässlichkeitsfrage werden Antwortkategorien vorgegeben wie z.B.:

- *sehr genau*
- *ziemlich genau*
- *eher ungenau*
- *grob geschätzt*

Bei der Zuordnung zu einer der beiden letzten Kategorien sollte dann nachgefragt werden: „*Und warum ist Ihre Antwort ungenau/grob?*“, um den Grund für die Ungenauigkeit zu erfahren.

Im Folgenden ist ein Beispiel dargestellt, bei dem die Qualität einer Rückerinnerungsfrage mittels der „Confidence Rating“-Technik überprüft wird.

Beispiel für die Überprüfung der Verlässlichkeit einer Antwort

Fragetext:

„Wie oft hat Ihre Mutter den Gottesdienst besucht als Sie 10 oder 11 Jahre alt waren?“

Skala auf Liste: mehr als einmal in der Woche/einmal in der Woche/ein- bis dreimal im Monat/mehrmals im Jahr/selten/nie (“kann ich nicht sagen“-Kategorie als Zusatzkategorie nur für den Interviewer sichtbar im Fragebogen)

(Vermutetes Problem: Die Rückerinnerung ist über einen derart großen Zeitraum nicht verlässlich.)

Antwort einer Testperson:

„Mehrmals im Jahr.“

Nach der Antwort wird die Verlässlichkeitsfrage gestellt:

Und wie sicher sind Sie sich bei der Einstufung der Gottesdienstbesuche Ihrer Mutter?“

Skala: 1 (sehr sicher) bis 5 (sehr unsicher)

Antwort auf die Verlässlichkeitsfrage:

„Wert 3“

Weitere Nachfassfrage (zu stellen bei den Skalenwerten 3, 4 und 5)

„Und warum sind Sie sich nicht sicher?“

Antwort auf die Nachfassfrage:

„Das ist zu lange her. Ich weiß es nicht mehr genau.“

Bewertung:

Die Verlässlichkeitstechnik in Kombination mit der Nachfassfrage zeigt, dass Rückerinnerungsfragen über einen so großen Zeitraum schwer zu erinnern sind und nicht immer valide Daten liefern. Ohne die Nachfassfragen wäre das Problem der Unsicherheit bei der Rückerinnerung nicht erkannt worden, da die Testperson formal korrekt geantwortet hat und dabei nicht von sich aus auf ihre Unsicherheit hingewiesen hat.

Empfehlung:

Die Kategorie „weiß nicht/ kann mich nicht mehr daran erinnern“ auf der Liste zusätzlich mit vorgeben.

Unsere Empfehlung für den Einsatz dieser Technik:

Die Technik ist unter Anwendungs- und Auswertungsgesichtspunkten einfach zu handhaben. Ihr Nachteil ist, dass man das eigentliche Problem, das einer wenig verlässlichen Antwort zugrunde liegt, nicht direkt erkennt. Aus diesem Grund ist die Technik nur in Kombination mit weiteren Nachfragen sinnvoll, d.h., wenn die Testperson angibt, bei der Beantwortung der Frage eher unsicher gewesen zu sein, muss nach dem Grund der Unsicherheit gefragt werden.

3.3 Paraphrasieren (Paraphrasing)

Beim Paraphrasieren sollen Testpersonen nach der Beantwortung der zu überprüfenden Frage den Fragetext in eigenen Worten wiederholen. Dabei ist darauf zu achten, dass die Testperson nicht versucht, sich wörtlich an den Text zu erinnern. Die Umsetzung in eigene Worte ist wichtig, weil der Forscher nur so überprüfen kann, ob die Testperson die Frage so wie vorgesehen interpretiert hat. Eine wörtliche Wiederholung des Textes würde lediglich eine gute Erinnerungsleistung der Testperson an den Fragetext dokumentieren.

Im Folgenden sind aus einem von ZUMA durchgeführten kognitiven Pretest beispielhaft sechs Paraphrasierungen einer Frage aufgeführt:

Beispiele für Paraphrasierungen

Frage:

„Wie stark wechseln die Wochentage, an denen Sie arbeiten, von Woche zu Woche?“

Skala: überhaupt nicht (1); kaum (2); ziemlich (3); sehr stark (4)

(Vermutetes Problem: Fragetext verwirrend.)

Testperson 1:

Die Testperson beantwortete die Frage mit Skalenwert 1 und paraphrasierte den Text folgendermaßen:

„Es ging um Wochentage“.

Testperson 2:

Die Testperson beantwortete die Frage mit Skalenwert 2 und paraphrasierte den Text folgendermaßen:

„Wie stark schwankt die Arbeitszeit am Ende des Tages.“

Testperson 3

Die Testperson beantwortete die Frage mit Skalenwert 1 und paraphrasierte den Text folgendermaßen:

„Ob ich in der Woche immer an den gleichen Tagen arbeite oder ob die Tage wechseln.“

Testperson 4

Die Testperson beantwortete die Frage mit Skalenwert 2 und paraphrasierte den Text folgendermaßen:

„Wechselt die Arbeitszeit überhaupt nicht, kaum, wenig oder sehr stark.“

Testperson 5

Die Testperson beantwortete die Frage mit Skalenwert 1 und paraphrasierte den Text folgendermaßen:

„Was? Die von eben? Oh! Das weiß ich gar nicht mehr. Wie oft schwanken die Wochentage. Das ist ja wie in der Schule.“

Testperson 6

Die Testperson beantwortete die Frage mit Skalenwert 4 und paraphrasierte den Text folgendermaßen:

„Meine Arbeitszeit geht von morgens 7 Uhr bis meistens 16.30 Uhr.“

Befund:

Mit Ausnahme von Testperson 3 zeigen die Paraphrasierungen, dass die wesentlichen Stimuli des Fragetextes nicht wiedergegeben werden können. Vermutlich ist der Fragetext zu kompliziert. Eine Veränderung des Textes wäre auf Grund dieser Ergebnisse zu empfehlen. Inwieweit die Testpersonen die Frage korrekt beantwortet haben, kann mit der Technik des Paraphrasierens nicht direkt erkannt werden, z.B. bringt Testperson 6 mit der Vergabe von Skalenwert „4“ zum Ausdruck, dass ihre Arbeitstage von Woche zu Woche sehr stark wechseln. Ihre Paraphrasierung jedoch bezieht sich auf ihre tägliche Arbeitszeit. Die Technik liefert hier nicht die gewünschte Information. Ob ihre Wochentage in der Tat stark wechseln (Skalenwert 4), müssen zusätzliche Nachfragen klären.

Empfehlung:

Frage teilen: „Arbeiten Sie grundsätzlich jede Woche an den gleichen Wochentagen?“ Falls ja: Weiter mit nächster Frage, falls nein: „Wie stark wechseln die Wochentage, an denen Sie arbeiten: Kaum, ziemlich oder sehr stark?“

Die neue Eingangsfrage ist für alle Befragten leichter zu verstehen.

Unsere Empfehlung für den Einsatz der Paraphrasierungstechnik:

Die Technik liefert in einem ersten Schritt lediglich Hinweise, in welche Richtung Befragte gedacht bzw. welche Aspekte sie mit einem Fragetext assoziiert haben. Sie liefert nicht unbedingt verlässliche Informationen über eventuelle Unsicherheiten beim Beantworten einer Frage. Denn: Auch wenn eine Testperson einen Fragetext korrekt wiedergibt, kann sie bei der Beantwortung der Frage Probleme gehabt haben. Solche Probleme können z.B. sein:

- *Jemand kann sich nicht mehr genau an Zeitpunkte, die in der Frage angesprochen werden erinnern.*
- *Jemand besitzt nicht die nötigen Informationen, um eine Frage zu beantworten.*
- *Jemand ist unsicher in Bezug auf die Bedeutung eines bestimmten Begriffs (z.B.: Welche Bedeutung hat der Begriff „Familie“? Ist damit die Kernfamilie oder die Großfamilie gemeint?)*
- *Jemand hat nicht verstanden, wie die Antwortskala benutzt werden soll.*

Setzt man die Technik des Paraphrasierens ein, so darf man bei einem korrekt wiedergegebenen Text nicht gleichzeitig auch von einem korrekten Frageverständnis ausgehen. Zuverlässige Informationen zum Frageverständnis kann die Technik nicht liefern, denn: Die Testperson wird bei dieser Technik nicht explizit aufgefordert, über ihre Unsicherheiten/Probleme zu sprechen, sondern sie wird lediglich gebeten, den Fragetext in eigenen Worten zu wiederholen.

Deshalb unsere Empfehlung: Nach der Technik des Paraphrasierens immer weitere Nachfragen einsetzen, die gezielt das Verständnis hinterfragen.

3.4 Sortiertechnik (Card Sorting)

Die Sortiertechnik eignet sich besonders dafür, um zu explorieren, wie Befragte Begriffe definieren. Dafür müssen im Vorfeld Situationen/Aussagen formuliert werden, die von den Testpersonen hinsichtlich dieses Begriffs zu bewerten sind. Jede Situation/Aussage wird auf eine Karte geschrieben und der jeweiligen Testperson vorgelegt. Dabei sollen die Karten so sortiert werden, wie sie nach Meinung der Testperson hinsichtlich des zu untersuchenden Begriffs zusammengehören.

Hier ein Beispiel für die Sortiertechnik, bei dem es um den Begriff „Verkehrsunfall“ geht. Mit Hilfe der Technik soll in Erfahrung gebracht werden, was Befragte unter einem „Verkehrsunfall“ verstehen.

Beispiel für die Sortiertechnik

Auf diesen Kärtchen sind verschiedene Situationen beschrieben, die sich auf einer Straße abspielen. Bitte sortieren Sie die Karten danach, ob Sie die beschriebene Situation als „Verkehrsunfall“ bezeichnen würden oder nicht.

(Testperson soll Karten auf zwei Felder sortieren: Ja, ist ein Verkehrsunfall/Nein, ist kein Verkehrs-unfall)

- **Zwei Autos stoßen auf einer Kreuzung zusammen.**
- **Ein Radfahrer gerät ohne Fremdverschulden mit seinem Fahrrad in eine Straßenbahnschiene und stürzt.**
- **Ein Bus-Fahrgast stürzt beim Aussteigen und bricht sich ein Bein.**
- **Ein Fußgänger rutscht auf dem Gehweg auf einer Bananenschale aus und verletzt sich.**

Je nach Zuordnung kann der Forscher dann feststellen, wie eng bzw. weit ein Verkehrsunfall definiert wird und kann dann diese Information bei der Fragenkonstruktion berücksichtigen.

Unsere Empfehlung für den Einsatz der Sortiertechnik:

Sortiertechniken arbeiten im Großen und Ganzen verlässlich. Für die Testpersonen ist der Umgang mit der Technik eher amüsant und abwechslungsreich. Für den Forscher stellt die Sortiertechnik insofern eine Anforderung dar, als er vor dem Test adäquate Kategorien konstruieren muss, die alle inhaltlichen Aspekte des zu untersuchenden Begriffs gut abdecken (s. Beispiel „Verkehrsunfall“).

3.5 Technik des lauten Denkens (Think Aloud)

Die Technik des lauten Denkens wurde ursprünglich von Psychologen zur möglichst lückenlosen Offenlegung gedanklicher Prozesse bei Problemlösungsaufgaben verwendet. Dieser Anspruch an eine Offenlegung möglichst aller Gedankenabläufe wurde beim Einsatz der Technik in der sozialwissenschaftlichen Umfrageforschung – seit den 90-iger Jahren – zurückgenommen, denn es hat sich gezeigt, dass Befragte bei der Beantwortung von Survey-Fragen häufig nicht in der Lage sind, ihre Gedankenabläufe zu formulieren bzw. diese überhaupt bewusst zu erkennen. Die Technik funktioniert am besten bei retrospektiven Fragen, weil Personen eher in der Lage sind, Rückerinnerungsprozesse zu formulieren als zum Beispiel die Überlegungen bei der Beantwortung einer Frage mittels eines Skalenwertes in Worte zu fassen.

Wie geht man bei dieser Technik vor?

Der Testleiter liest zuerst die zu überprüfende Frage vor und bittet dann die Testperson, alle Überlegungen, die zur Beantwortung der Frage führen zu formulieren. Der Testleiter soll bei entstehenden Pausen nur insofern helfend eingreifen, als er die Testperson noch einmal zum lauten Denken auffordert. Die Äußerungen der Testperson müssen – am besten videogestützt – aufgezeichnet werden. Die aufgezeichneten Formulierungen werden in der Literatur auch als „verbal protocols“ bezeichnet (siehe Ericsson/Simon, 1984).

Das laute Denken muss mit der Testperson zunächst geübt werden. Eine Übungsfrage kann z.B. sein:

“Stellen Sie sich bitte die Wohnung oder das Haus vor, in der/in dem Sie leben. Wie viele Fenster gibt es dort? Bitte sagen Sie mir, was Sie sehen und denken, während Sie die Fenster zählen.”

Die Testperson erfüllt die Aufgabe dann korrekt, wenn sie beim Zusammenzählen der Fenster möglichst kontinuierlich spricht und dem Testleiter somit vermittelt, wie sie im Detail vorgeht oder auch mit evtl. auftretenden Unklarheiten umgeht (z.B.: Ist ein Fenster bestehend aus mehreren Fensterrahmen als **ein** Fenster zu werten? Oder: Sollen auch Kellerfenster dazu gezählt werden?)

An den folgenden Beispielen aus kognitiven Tests soll gezeigt werden, wie Testpersonen mit der Technik umgehen und welchen Nutzen die Technik dem Forscher bringt.

Beispiel 1: Methode funktioniert

Testperson formuliert relevante Überlegungen, die zur Beantwortung der Frage führen

Frage:

“Wie oft waren Sie in den letzten 12 Monaten beim Zahnarzt?”

(Vermutetes Problem: Sind die Testpersonen in der Lage, sich zu erinnern, bzw. wenn ja, wie genau können sie sich erinnern?)

Testperson:

„Oh, lassen Sie mich nachdenken. Ich glaube, ich war ein Mal zum Nachsehen und um die Zähne reinigen zu lassen. An ein anderes Mal kann ich mich nicht erinnern. Ich bin mir nicht sicher. Aber irgendwann um diese Zeit habe ich Probleme mit meinem Kiefer gehabt, aber sie konnten nichts finden. Ich bin mir aber nicht sicher, welcher Monat das war. Ich würde also sagen ein Mal.“

Befund:

Die Methode hat funktioniert: Die Testperson formuliert informative Überlegungen. Sie versucht, sich an einzelne Besuche zu erinnern und nicht eine Schätzung über den Zeitraum abzugeben. Durch die Ausführungen wird jedoch auch deutlich, dass sich die Testperson nicht zuverlässig erinnern kann. Der Zeitraum von 12 Monaten ist offensichtlich zu groß.

Beispiel 2: Methode funktioniert nicht

Testperson formuliert überhaupt keine bzw. keine eindeutig relevanten Überlegungen

Frage:

“Wenn Sie einmal an die Tage denken, an denen Sie fernsehen: Wie lange – ich meine in Stunden und Minuten – sehen Sie da im allgemeinen an einem Tag fern?”

(Vermutetes Problem: Sind die Testpersonen in der Lage, eine durchschnittliche Stundenzahl zu berechnen.)

Testperson 1:

„Drei Stunden.“

Befund:

Die Methode hat nicht funktioniert: Die Testperson formuliert nicht, welche Überlegungen zu ihrer Antwort führten, sondern gibt sofort die Antwort. Hinweise auf mögliche Probleme bei der Beantwortung der Frage sind nicht zu erkennen.

Testperson 2:

„Vor der Tagesschau mache ich den Fernseher selten an. Nach der Tagesschau schaue ich Filme oder Diskussionen. Bei schönem Wetter schaue ich überhaupt nicht fern.“

Befund:

Die Testperson hat zwar Gedanken formuliert, die Ausführungen sind jedoch in Bezug auf die eigentliche Fragestellung nicht eindeutig. Sie liefern keine Informationen darüber, ob die Testperson in der Lage gewesen wäre, einen Durchschnittswert zu bilden.

Unsere Empfehlung für den Einsatz der Technik des lauten Denkens:

Wir empfehlen, die Technik nur bei retrospektiven Fragen oder Fragen, die eine Problemlösung verlangen einzusetzen, weil Erinnerungs- und Problemlösungsprozesse noch am ehesten formuliert werden können.

Grundsätzlich gilt aber:

1. Das laute Denken fällt den meisten Testpersonen schwer - manche können es überhaupt nicht leisten. Nach unseren Erfahrungen sind nur ca. 50% der Testpersonen in der Lage, ihre Gedanken laut zu formulieren.
2. Auch wenn Testpersonen laut denken, enthalten diese Ausführungen nicht immer brauchbare Informationen, weil diese für den zu überprüfenden Fragenaspekt irrelevant sind. Die Methode produziert in solchen Fällen lediglich „unbrauchbares Material“.
3. Die Übungsphase ist aufwendig und zeitintensiv; sie muss vor jedem Testinterview durchgeführt werden.
4. Auch die Auswertungsphase ist zeitintensiv. Da die Technik eine „freie Rede“ zulässt, d.h. die Ausführungen der Testperson nicht reglementiert sind, werden in manchen Fällen große Textmengen produziert, die aber nur zum Teil Informationen zum überprüften Sachverhalt/vermuteten Problem enthalten.
5. Das laute Denken selbst kann die Situation dahingehend verändern, als Testpersonen „künstlich“ reagieren und die Äußerungen unrealistisch werden.

Fazit:

Die Technik des lauten Denkens produziert - wenn überhaupt - eher unzuverlässige Informationen. Darüber hinaus ist der Aufwand für den Einsatz der Technik unseres Erachtens im Verhältnis zum „Output“ zu hoch.

4. Empfehlungen zur Durchführung kognitiver Interviews

Bei der Durchführung kognitiver Interviews sollten folgende Punkte beachtet werden:

- Idealerweise sollten kognitive Interviews in einem kognitiven Labor stattfinden, wo nicht nur die Möglichkeit der Videoaufzeichnung besteht, sondern das Interview auch in einem angrenzenden Raum am Monitor verfolgt werden kann. Durch die Videoaufzeichnung können die Interviews später nicht nur akustisch, sondern auch im Hinblick auf visuelle Auffälligkeiten (z.B. Mimik, Gestik) analysiert werden. Ein kognitives Labor ist jedoch für die Durchführung kognitiver Interviews nicht unbedingt nötig. Es genügt auch ein normaler Büroraum. Jedes Interview sollte jedoch zumindest mit Tonband aufgezeichnet werden, vor allem deshalb, weil sich der Testleiter intensiv auf die Interviewführung konzentrieren muss und deshalb nicht ausführlich mitprotokollieren kann.
- Im Hinblick auf die Anzahl der Interviews im Rahmen eines Tests können wir keine klare Empfehlung aussprechen. Bei ZUMA werden in der Regel zwischen 5 und 15 Interviews durchgeführt.
- Ein kognitives Interview sollte die Dauer von 90 Minuten nicht überschreiten. Nach 90 Minuten lässt erfahrungsgemäß die Konzentration und das Interesse bei den Testpersonen deutlich nach.

- Die Testpersonen sind nicht beliebig, sondern gezielt nach Merkmalen zu rekrutieren. Neben den üblichen demografischen Merkmalen wie Alter, Geschlecht und Bildung muss überlegt werden, ob die Testpersonen weitere Merkmale aufweisen müssen, um die Inhalte der zu überprüfenden Fragen sinnvoll testen zu können (z.B. können Fragen zur Internetnutzung im Schulunterricht nur an Lehrern/Schülern getestet werden, die sich damit befassen und damit den Personen entsprechen, die später befragt werden sollen)
- Die Testpersonen werden vor dem kognitiven Interview über dessen Testcharakter informiert und gebeten, ihre Meinung und ihre Probleme bei den einzelnen Fragen zu formulieren und bei der Verbesserung der Fragen mitzuwirken. Dafür erhalten die Testpersonen ein Honorar, das normalerweise bei ca. € 25.- pro einstündigem Interview liegt – Wegezeiten und Fahrtkosten eingeschlossen.
- Wichtig ist der Hinweis an die Testpersonen, dass nicht sie selbst getestet werden, sondern ein Fragebogen bzw. einzelne Fragen.
- Zur Durchführung kognitiver Interviews gibt es keine festen Regeln. Die Spannweite reicht von einer völlig offenen bis zu einer stark standardisierten Vorgehensweise. Völlig offen bedeutet in diesem Zusammenhang: Die Techniken bzw. Nachfragen des Testleiters sind vor dem Interview nicht festgelegt. Standardisiert bedeutet: Alle Techniken bzw. Nachfragen sind dem Testleiter fest vorgegeben. Beide Extreme haben sich in der Praxis wenig bewährt. Eine völlig offene Vorgehensweise liefert durch unterschiedlich gesetzte Stimuli wenig vergleichbares Material; eine stark standardisierte Vorgehensweise produziert unter Umständen weniger Informationen, weil der Testleiter nicht spontan und frei auf die Äußerungen der Testpersonen reagieren kann. Aus diesen Gründen schlagen wir zur Durchführung kognitiver Interviews eine Kombination einer standardisierten und einer offenen Vorgehensweise vor. Wir empfehlen, die Techniken bzw. Nachfragen für jede zu testende Frage vor der Testphase in einem so genannten „Evaluationsfragebogen“ festzulegen, den der Testleiter später im kognitiven Interview weitgehend standardisiert erhebt. Bei Bedarf kann der Testleiter jedoch auch spontan reagieren und zusätzliche nicht vorbereitete Fragen stellen oder kognitive Techniken anwenden.
- Wenn in einer frühen Phase der Fragenentwicklung technisch noch unausgereifte Frage-Versionen besonders bezüglich ihres Verständnisses getestet werden sollen, spielt es noch keine wesentliche Rolle, in welchem Modus diese später tatsächlich erhoben werden. Sollten jedoch Fragen bereits im Modus bzw. Format der späteren Erhebung vorliegen, dann ist es ratsam, sie auch den Testpersonen in diesem Modus zu präsentieren, z.B. sollten Fragen aus einem Paper-and-Pencil-Selbstaussfüllerfragebogen den Testpersonen auch als Paper-and-Pencil-Selbstaussfüllerfragen vorgelegt werden, Fragen aus einem Online-Fragebogen wiederum sollten Testpersonen am Computer beantworten.
- Falls genügend Ressourcen (vor allem Zeit) vorhanden ist, empfehlen wir, die auf Grund der Ergebnisse kognitiver Interviews veränderten Fragen einem weiteren kognitiven Pretest zu unterziehen, um mit dieser iterativen Vorgehensweise vor allem die Wirksamkeit der Veränderungen zu evaluieren.

5. Wie wertet man kognitive Interviews aus?

Wir empfehlen, bei der Auswertung kognitiver Interviews für jede getestete Frage eine fallspezifische Auflistung⁶ aller Befragtenäußerungen zu erstellen. Diese Auflistungen enthalten:

- Die Antworten auf die zu testende Frage,
- spontane Reaktionen der Testpersonen (Zusatzäußerungen, die nicht als Antwort auf die zu testende Frage zu werten sind),
- die Antworten bzw. Reaktionen auf die kognitiven Nachfragen bzw. Techniken,
- Anmerkungen des Testleiters.

Die Antworten auf die kognitiven Techniken müssen fallspezifisch dahingehend analysiert bzw. kategorisiert werden, inwieweit sie über das Verständnis bzw. das „Funktionieren“ der getesteten Frage Aufschluss geben. Zusätzlich können auch die spontanen Reaktionen der Testpersonen sowie ihre Antworten auf die zu testende Frage Hinweise geben. Dieser Arbeitsschritt ist weitgehend vergleichbar mit der Auswertung von Antworten auf offene Fragen.

6. Finanzieller und zeitlicher Aufwand

Finanzieller Aufwand:

Werden die Entwicklungsarbeiten und die Interviews vom Projektteam selbst durchgeführt, so fallen als Kosten lediglich die Befragtenhonorare an – vorausgesetzt, das Team verfügt über eine adäquate räumliche und technische Ausstattung.

Zeitlicher Aufwand:

Der zeitliche Aufwand ist hoch. Dies gilt für alle Phasen eines kognitiven Pretests.

- In der Vorphase muss ausreichend Zeit zur Rekrutierung der Testpersonen vorgesehen werden. Je spezifischer die Merkmale der Testpersonen sind, desto höher ist dabei der zeitliche Aufwand (gesucht werden z.B. Personen, die einer bestimmten Altersgruppe und gleichzeitig einer bestimmten Berufsgruppe angehören).
- Für die Festlegung der kognitiven Techniken und deren Umsetzung in Zusatzfragen werden in der Regel ca. 1,5 bis 2 Wochen benötigt. Dieser Zeitaufwand hängt jedoch stark vom Diskussionsbedarf des Projektteams und von der Anzahl der zu testenden Fragen ab.
- Wenn die Interviews von zwei Testleitern durchgeführt werden, so kann man bei der Durchführung von 15 bis 20 Interviews als Richtwert ca. 1,5 bis 2 Wochen festsetzen.
- Für die Auswertung und Dokumentation ist der zeitliche Aufwand schwer einzuschätzen. Er hängt vor allem davon ab, wie intensiv das erhobene Datenmaterial (z.B. spontane Reaktionen der Testpersonen, Antworten der Testpersonen auf die zu testenden Fragen und auf die kognitiven Nachfragen) ausgewertet bzw. bearbeitet wird. Z.B. ist die Transkription aller Pretestinterviews als Basis für die Ergebnisdarstellung zwar wünschenswert, jedoch im Zeitplan der meisten Projekte nicht durchführbar. Für eine fall- und fragenspezifische Auswertung von ca. 15 Interviews (Dauer eine Stunde pro Fall) muss man ganz grob mit einer Bearbeitungszeit von mindestens einer Woche rechnen.

⁶ Ein Beispiel für ein Auflistungsschema befindet sich im Anhang.

Dass mit der Durchführung kognitiver Interviews ein höherer Aufwand an zeitlichen Ressourcen verbunden ist als mit der Durchführung eines Beobachtungs-Pretests, liegt auf der Hand. Der Zuwachs an wichtigen Informationen über die Qualität einer Frage ist jedoch so hoch, dass wir Umfrageforschern die Durchführung kognitiver Interviews bei der Entwicklung eines Fragebogens auf jeden Fall empfehlen.

Dabei möchten wir abschließend noch auf zwei Punkte hinweisen:

- Kognitive Interviews allein führen nicht zu verbesserten Fragen, sondern erst die Umsetzung bzw. Verarbeitung der Ergebnisse. Kognitive Interviews können lediglich potentielle Probleme identifizieren.
- Der nach Abschluss der kognitiven Testphase fertig gestellte Fragebogen muss abschließend, d.h. vor seinem endgültigen Einsatz mit anderen Pretestmethoden auf seine Funktionsfähigkeit geprüft werden. Hier bietet sich in erster Linie der klassische Standard- oder Beobachtungs-Pretest an, der mit Techniken wie „Behaviour Coding“, „Interviewer-Debriefing“ oder „Befragten-Debriefing“⁷ arbeitet. Mit solchen Techniken können vor allem Aspekte, die den gesamten Fragebogen betreffen, wie z.B. Probleme der Interviewer mit der Handhabung des Fragebogens, die Wirkung der Thematik oder die Akzeptanz der Befragungsdauer evaluiert werden. Dabei ist es unbedingt erforderlich, dass der Fragebogen in einer Art „Generalprobe“ so zum Einsatz kommt, wie er später in der Hauptstudie erhoben wird.

Literatur:

Belson, W. A., 1981: *The Design and Understanding of Survey Questions*. Aldershot, England: Gower.

Cantril, H./Fried, E., 1944: *The Meaning of Questions*. In Cantril, H. (Ed.): *Gauging Public Opinion*. Princeton: Princeton University Press.

DeMaio, T. J./Rothgeb, J. M., 1996: *Cognitive Interviewing Techniques: In the Lab and in the Field*. In Schwarz, N. & Sudman, S. (Eds.). *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*. San Francisco: Jossey-Bass Publishers, 177-196.

Ericsson, K.A./Simon, H.A., 1984: *Protocol Analysis: Verbal Reports as Data*. Cambridge: MIT Press.

Faulbaum, F./Prüfer, P./Rexroth, M., 2009: *Was ist eine gute Frage? Die systematische Evaluation der Fragenqualität*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Foddy, W., 1993: *Constructing Questions for Interviews and Questionnaires*. Cambridge: Cambridge University Press.

Foddy, W., 1998: *An empirical Evaluation of In-depth probes used to pretest Survey Questions*. S. 103-133 in: *Sociological Methods and Research*, 27, 1998.

Fowler, F. J., 1995: *Improving Survey Questions*. Thousand Oaks: Sage.

⁷ Eine Beschreibung dieser Techniken ist nachzulesen in Prüfer/Rexroth 1996.

- Jabine, T./ Straf, M./ Tanur, J./ Tourangeau, R. (Eds.), 1984: *Cognitive Aspects of Survey Design: Building a Bridge between Disciplines*. Washington, DC: National Academy Press.
- Presser, S. et al., 2004: *Methods for Testing and Evaluating Survey Questions*. New York: Wiley.
- Prüfer, P./Rexroth, M., 1996: *Verfahren zur Evaluation von Survey-Fragen: Ein Überblick*. ZUMA-Arbeitsbericht 96/05.
- Prüfer, P./Rexroth, M., 2000: *Zwei-Phasen-Pretesting*. ZUMA-Arbeitsbericht 2000/08.
- Prüfer, P./Rexroth, M./Fowler, F.J. (Eds.), 2003: *ZUMA Nachrichten Spezial 9: QUEST 2003. Proceedings of the 4th Conference on Questionnaire Evaluation Standards*.
- Schuman, H., 1966: *The Random Probe: A Technique for Evaluating the Validity of Closed Questions*. *American Sociological Review* 31: 218-222.
- Sirken, M.G. et al. (Eds.), 1999: *Cognition and Survey Research*. New York: John Wiley & Sons.
- Sudman, S./Bradburn, N./Schwarz, N., 1996: *Thinking about Answers. The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass.
- Tanur, J.M. (Ed.), 1992: *Questions about Questions*. New York: Russell Sage Foundation.
- Tourangeau, R., Rips, L. J., Rasinski, K., 2000: *The Psychology of Survey Response*. Cambridge University Press.
- Willis, G.B., 2005: *Cognitive Interviewing. A Tool for Improving Questionnaire Design*. Thousand Oaks: Sage Publications.
- Willis, G.B./ Lessler, J., 1999: *Question Appraisal System, QUAS-99*. Research Triangle Institute.

**Beispiel für ein Auflistungsschema zur Auswertung einer Frage über alle Interviews
(10 kognitive Interviews)**

| Inter- view Nr. | Antworten auf die zu testende Frage (Skalenwert/Antwort- kategorie etc.) | Spontane Reaktionen der Testpersonen | Antworten/Reaktionen auf die kognitiven Fragen/Techniken | Anmerkungen des Testleiters |
|--------------------------------|---|---|---|------------------------------------|
| 1 | | | | |
| 2 | | | | |
| 3 | | | | |
| 4 | | | | |
| 5 | | | | |
| 6 | | | | |
| 7 | | | | |
| 8 | | | | |
| 9 | | | | |
| 10 | | | | |