

GESIS Survey Guidelines

Kognitives Pretesting

Timo Lenzner, Cornelia Neuert & Wanda Otto

Zusammenfassung

Die Durchführung eines Pretests gilt in der Umfrageforschung gemeinhin als unabdingbare Voraussetzung für die erfolgreiche Entwicklung eines Fragebogens und damit für die Qualität der in einer Befragung erhobenen Daten. Dieser Beitrag bietet eine Einführung in das Thema Fragebogen-Pretesting und beschäftigt sich insbesondere mit den zentralen Fragen der Planung und Durchführung von kognitiven Pretests. Im Einzelnen werden folgende Fragen behandelt: Was ist ein Pretest und warum sollte man Fragebogen-Pretests durchführen? Welche Pretestingverfahren stehen zur Verfügung und für welches sollte ich mich entscheiden? Welche Techniken werden in kognitiven Pretests angewandt? Wie führt man kognitive Pretests durch? Wie wertet man kognitive Pretests aus?

Zitierung

Lenzner, Timo, Neuert, Cornelia und Otto, Wanda (2015). Kognitives Pretesting. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS Survey Guidelines). DOI: 10.15465/gesis-sg_010

Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung - Nicht kommerziell 4.0 International Lizenz (CC BY-NC).



1. Was ist ein Pretest und warum sollte man Fragebogen-Pretests durchführen?

Wenn man von einem Fragebogen-Pretest spricht, meint man damit ganz allgemein die Evaluation bzw. das Testen eines Fragebogens vor dessen Einsatz in der eigentlichen Erhebung. Pretests sind damit ein wesentlicher Bestandteil der Fragebogenentwicklung. Ziel und Zweck eines Pretest ist es, u.a. Auskunft zu geben über (vgl. Converse & Presser, 1986; Porst, 2000):

- die Verständlichkeit der Fragen
Stimmt die von Befragten assoziierte Bedeutung einer Frage mit der vom Forscher intendierten Bedeutung überein? Interpretieren die Befragten die Bedeutung einer Frage in der gleichen Weise?
- Probleme der Befragungsperson mit ihrer Aufgabe
Wie schwer fällt es den Befragten, die Frage zu verstehen und zu beantworten? Ist der Sachverhalt, der in der Frage angesprochen wird, eventuell unbekannt oder heikel?
- Interesse und Aufmerksamkeit des Befragten bei einzelnen Fragen
Setzen im Laufe der Befragung Ermüdungseffekte ein? Sind aus Sicht des Befragten (einzelne) Fragen redundant?
- Häufigkeitsverteilungen der Antworten
Wird die volle Skalenbreite ausgenutzt?
- Kontexteffekte und Probleme mit der Reihenfolge der Fragen
Beeinflussen vorangegangene Fragen die Beantwortung nachfolgender Fragen?
- Probleme des Interviewers
Können Interviewer klar erkennen, was sie vorlesen sollen und was nicht?
- Technische Probleme mit Fragebogen und Befragungshilfen (z.B. Listen, Kärtchen)
- die Zeitdauer der Befragung

Alle diese Aspekte liefern wichtige Hinweise bezüglich der Funktionsfähigkeit eines Fragebogens und damit auf die Qualität der mit dem Fragebogen erhobenen Daten. Eine Überprüfung des Fragebogens am Schreibtisch ist generell nicht ausreichend, um „gute“ Daten sicherzustellen, denn wie bereits Sudman und Bradburn (1982: 283) bemerken: „Even after years of experience, no expert can write a perfect questionnaire.“ Nur mit Hilfe von empirischen Pretests lässt sich überprüfen, ob die Fragen eines Fragebogens tatsächlich das messen, was sie messen sollen und ob sie reliable und valide Antworten liefern.

2. Welche Pretestingverfahren stehen zur Verfügung und für welches sollte ich mich entscheiden?

Zur Evaluation eines Fragebogens stehen Forschern eine Reihe ganz unterschiedlicher Pretestingverfahren zur Verfügung, wie z.B. Standard-Pretest, Kognitives Interview, Behavior Coding, Befragten-Debriefing, Gruppendiskussion, Expert Review, Eyetracking oder Web Probing (Behr, Braun, Kaczmirek, & Bandilla, 2013; Campanelli 2008; Faulbaum, Prüfer, & Rexroth, 2009; Prüfer & Rexroth, 1996:). Jedes dieser Verfahren weist ganz spezifische Stärken und Schwächen auf. Die Entscheidung für ein oder mehrere Verfahren hängt entscheidend von dem jeweiligen Anwendungsfeld und Erkenntnisinteresse ab. Im Folgenden konzentrieren wir uns auf zwei Pretestingverfahren, den Standard-Pretest, der vor jeder Erhebung zwingend durchgeführt werden sollte und den kognitiven Pretest, der im Idealfall zusätzlich zum Standard-Pretest und diesem vorgeschaltet, als Teil der Fragebogenentwicklung eingesetzt werden sollte.

2.1 Der Standard-Pretest

Beim Standard-Pretest wird der Fragebogen unter möglichst realistischen Bedingungen der Hauptstudie eingesetzt. Es handelt sich dabei also gewissermaßen um eine Simulation der Haupterhebung mit einer Stichprobe von etwa 10 bis 200 Personen (Prüfer & Rexroth, 1996). Der Standard-Pretest sollte im selben Befragungsmodus (persönlich-mündlich, telefonisch, schriftlich, online, etc.) durchgeführt werden, der auch in der Haupterhebung zum Einsatz kommt. In der Regel werden die Befragten nicht über den Testcharakter der Befragung informiert, um den Beantwortungsprozess dadurch nicht zu beeinflussen.

Ziel des Standard-Pretests ist es, den gesamten Ablauf eines Interviews bezüglich Praktikabilität bzw. den gesamten Fragebogen auf Handhabbarkeit zu überprüfen. Damit erhält man in der Regel verlässliche Informationen über technische Mängel des Fragebogens (z.B. Filterführungen sind nicht korrekt, Interviewer haben Probleme mit der Handhabung des Fragebogens), Häufigkeitsverteilungen der Antworten und die durchschnittliche Befragungsdauer in der Haupterhebung. Darüber hinaus kann der Standard-Pretest auch begrenzte Informationen zum Frageverständnis liefern. Beispielsweise können im Rahmen von interviewer-administrierten Erhebungen die Interviewer dazu angehalten werden, Probleme und Auffälligkeiten auf Seiten der Befragten zu notieren und zu berichten. Im Allgemeinen handelt es sich beim Standard-Pretest jedoch um ein *passives Verfahren*, bei dem der Interviewer oder Pretester lediglich beobachtet, ohne das Verständnis von Fragen aktiv zu hinterfragen. In der Regel wird ein Verständnisproblem auf Seiten der Befragten bei einer Frage daher nur dann aufgedeckt, wenn dieses vom Befragten selbst angesprochen wird. Durch einen Standard-Pretest erhält man daher meist nur wenige, unsystematische und oberflächliche Informationen zum Frageverständnis.

2.2 Kognitive Pretests

Kognitive Pretests (oder kognitive Interviews) zählen zu den *aktiven Pretestverfahren*, da hierbei das Vorgehen der Befragten bei der Beantwortung von Fragen aktiv hinterfragt und untersucht wird (Beatty & Willis, 2007; Prüfer & Rexroth, 2005; Willis 2005). Kognitive Pretests werden meist in der Entwicklungsphase eines Fragebogens eingesetzt, um Einblick in die kognitiven Prozesse zu bekommen, die beim Beantworten von Fragen ablaufen:

- Wie interpretieren Befragte Fragen oder Begriffe?
- Wie rufen sie Informationen und Ereignisse aus dem Gedächtnis ab?
- Wie kommen sie zu einer Entscheidung darüber, wie sie antworten?

- Wie ordnen sie ihre „intern“ ermittelte Antwort formalen Antwortkategorien zu?

Das eigentliche Ziel dabei besteht darin, durch diese Informationen Hinweise auf unterschiedlichste Frage-Probleme zu erhalten. Kognitive Pretests eignen sich besonders dazu, die Verständlichkeit von Fragen zu überprüfen, Probleme der Befragten bei der Beantwortung des Fragebogens zu identifizieren, die Ursachen dieser Probleme aufzudecken und anhand der/dieser gewonnenen Erkenntnisse Verbesserungsvorschläge zu generieren.

Im Gegensatz zum Standard-Pretest steht bei kognitiven Pretests das Testen von Einzelfragen im Mittelpunkt und nicht die Evaluation des gesamten Fragebogens. Kognitive Pretests sind generell für alle Befragungsmodi geeignet, d.h. sie können unabhängig davon eingesetzt werden, ob die spätere Befragung persönlich-mündlich, telefonisch, schriftlich oder online erfolgt. Optimal für die Durchführung der Hauptstudie ist es, wenn der Fragebogen im Anschluss an einen kognitiven Pretest zusätzlich noch einen Standard-Pretest durchläuft, um in dessen Rahmen seine Funktionsfähigkeit unter realistischen Hauptstudien-Bedingungen zu testen.

3. Welche Techniken werden in kognitiven Pretests angewandt?

3.1 Die Technik des lauten Denkens (Think Aloud)

Beim Think Aloud werden die Befragungspersonen aufgefordert „laut zu denken“ und sämtliche Gedankengänge, die zur Antwort führen bzw. geführt haben, zu verbalisieren. Dadurch soll der Antwortprozess erkennbar werden und damit auch eventuell vorhandene Probleme mit dem Fragenverständnis. Das Think Aloud kann entweder während der Beantwortung einer Frage (concurrent think aloud) oder nach der Beantwortung einer Frage (retrospective think aloud) erfolgen. Bei einem concurrent think aloud würde man die zu testende Frage beispielsweise mit folgender Instruktion einleiten (Porst, 2014): „Bitte sagen Sie mir bei der folgenden Frage alles, an was Sie denken oder was Ihnen durch den Kopf geht, bevor Sie die Frage beantworten. Sagen Sie bitte auch Dinge, die Ihnen vielleicht unwichtig erscheinen. Die Frage lautet:“

3.2 Nachfragetechniken (Probing)

Beim Probing handelt es sich um eine Technik, bei der Begriffe, Fragetexte oder gegebene Antworten mittels einer oder mehrerer Zusatzfragen (probes) hinterfragt werden. Dadurch können zusätzliche Informationen über das Verständnis von Fragen gewonnen werden. Je nach Erkenntnisinteresse können verschiedene Probing-Varianten eingesetzt werden: Comprehension Probing (Nachfragen zum Verständnis, z.B. „Was verstehen Sie in dieser Frage unter ‚einer beruflichen Tätigkeit mit hoher Verantwortung?‘“), Category Selection Probing (Nachfragen zur Wahl der Antwortkategorie, z.B. „Können Sie Ihre Antwort bitte begründen?“), Information Retrieval Probing (Nachfragen zur Informationsbeschaffung, z.B. „Wie haben Sie sich daran erinnert, dass Sie in den letzten 12 Monaten [...] mal beim Arzt waren?“) oder General/Elaborative Probing (Unspezifisches Nachfragen, z.B. „Können Sie mir Ihre Antwort bitte noch etwas näher erläutern?“). Probing-Fragen können entweder direkt im Anschluss an die Beantwortung einer Survey-Frage gestellt werden oder erst nachdem die Testpersonen den gesamten Fragebogen beantwortet haben.

3.3 Paraphrasieren (Paraphrasing)

Beim Paraphrasieren soll die Befragungsperson nach Beantwortung einer Frage den Fragetext in eigenen Worten wiedergeben („Bitte wiederholen Sie die Frage, die ich Ihnen vorgelesen habe, noch einmal in Ihren eigenen Worten.“). Idealerweise liefert diese Verbalisierung Informationen darüber, ob bzw. wie Personen eine Frage verstanden haben und ob das Verständnis des Befragten mit dem des Forschers übereinstimmt. Dabei gilt zu beachten, dass man aufgrund einer mangelhaften Wiedergabeleistung nicht zwangsläufig auf ein fehlendes Fragenverständnis schließen kann. Generell eignet sich das Paraphrasieren nicht für kurze Faktfragen, da diese von den Befragten meist wörtlich erinnert werden können und man damit lediglich Informationen über die Gedächtnisleistung der Befragten erhält, nicht aber über ihr Verständnis der Frage.

3.4 Bewertung der Verlässlichkeit der Antwort (Confidence Rating)

Bei dieser Technik werden die Befragten gebeten, nach der eigentlichen Antwort den Grad der Verlässlichkeit ihrer Antwort zu bewerten und beispielsweise anzugeben, wie sicher sie sich sind, korrekt geantwortet zu haben („Wie sicher sind Sie, dass Sie in den letzten 12 Monaten [...] mal beim Arzt waren?“). Falls Befragte angeben, bei der Beantwortung einer Frage eher unsicher gewesen zu sein, sollte immer eine Nachfrage nach dem Grund für die Unsicherheit folgen.

3.5 Sortiertechniken (Sorting)

Sortiertechniken werden angewendet, um zu untersuchen, wie Testpersonen Begriffe oder Situationen bestimmten Kategorien zuordnen (z.B. welche Beschreibungen einer Unfallsituation Befragte als „Verkehrsunfall“ bezeichnen und welche nicht). Wir unterscheiden zwischen freiem und dimensionalem Sortieren:

- Beim freien Sortieren sollen Befragte vorgegebene Items nach eigenen Kriterien gruppieren
- Beim dimensionalen Sortieren sollen vorgegebene Items nach vorher festgelegten Kriterien sortiert werden.

Diese Technik stellt Forscher vor die Herausforderung, vor dem Test adäquate Kategorien zu konstruieren, die alle inhaltlichen Aspekte des zu untersuchenden Begriffs vollständig abdecken.

Bezüglich der Frage, wie abwechslungsreich der Einsatz der verschiedenen kognitiven Techniken innerhalb eines Pretests gestaltet werden sollte, existieren keine allgemeingültigen Regeln. Einerseits ist ein gewisses Maß an Abwechslung sicher hilfreich um die Befragten nicht zu ermüden und ihre Aufmerksamkeit nicht von den eigentlichen Fragen weg zu den erwarteten Nachfragen zu lenken. Andererseits sollte man nur um der Abwechslung willen keine Techniken einsetzen, die im Pretest möglicherweise gar keinen Erkenntnisgewinn erbringen. Der Einsatz einer kognitiven Technik sollte immer durch das Erkenntnisinteresse des Interviewers oder das Verhalten des Befragten bestimmt sein.

4. Wie führt man kognitive Pretests durch?

Zur Durchführung von kognitiven Interviews gibt es keine festen Regeln. Idealerweise finden sie in einem ruhigen, geschlossenen Raum statt und werden per Videokamera (notfalls auch lediglich per Diktiergerät) aufgezeichnet. Die Videoaufzeichnung hat den Vorteil, dass die Interviews später nicht nur akustisch, sondern zusätzlich auch im Hinblick auf visuelle Auffälligkeiten (Mimik, Gestik) analysiert werden können. Auf eine Aufzeichnung der Interviews (egal ob Audio oder Video) sollte generell nicht

verzichtet werden, da dies sowohl die Durchführung des Interviews als auch die Auswertung des Gesprächs vereinfacht und qualitativ verbessert.

Bezüglich der Anzahl der Interviews, die im Rahmen eines kognitiven Pretests durchgeführt werden sollten, existieren unterschiedliche Praktiken und Empfehlungen. In der Regel werden zwischen 5 und 30 Interviews pro Pretest(runde) durchgeführt, da die schwerwiegendsten Frage-Probleme meist bereits auf Basis einer relativ kleinen Anzahl von Interviews identifiziert werden (Willis, 2005). Andererseits konnten Blair und Conrad (2011) zeigen, dass mit einer zunehmenden Anzahl von Interviews auch die Wahrscheinlichkeit steigt, weitere bedeutsame Frage-Probleme zu entdecken. Berücksichtigt man die großen Datenmengen (verbale Textdaten), die im Rahmen von kognitiven Interviews produziert werden und die zur Auswertung interpretiert werden müssen, erscheint die Durchführung von mehr als 30 Interviews pro Pretest jedoch als recht impraktikabel. Aufgrund solcher Kosten-Nutzen-Abwägungen arbeitet man daher meist mit ca. 20 Personen pro kognitiven Pretest. Falls genügend Ressourcen zur Verfügung stehen, wird empfohlen, 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 (Prüfer & Rexroth, 2005; Willis, 2005).

Bei der Durchführung kognitiver Interviews reicht die Spannweite von (fast) völlig unstrukturiert bis (fast) völlig standardisiert. Beim völlig unstrukturierten Vorgehen setzt der Interviewer die kognitiven Techniken ausschließlich spontan und nur als Reaktion auf das Verhalten des Befragten ein. Beim völlig standardisierten Vorgehen werden die kognitiven Techniken im Voraus in einem Interviewleitfaden festgehalten und der Interviewer ist angehalten, den Leitfaden möglichst standardisiert zu administrieren. Wir empfehlen hier eine Mischung aus standardisiertem und unstandardisiertem Vorgehen. Im Vorfeld sollte ein Interviewleitfaden erstellt werden, der die zu testenden Fragen, die zugeordneten kognitiven Techniken und Platz für Kommentare des Befragten sowie Mitschriften und Kommentare des Interviewers enthält. Bei Bedarf können während des Interviews spontan weitere, nicht vorbereitete kognitive Techniken und Fragen nachgeschoben werden. Dabei sollte es den Interviewern überlassen sein, so lange nachzufragen, bis sie sicher sind, alle nötigen Informationen erhalten zu haben.

Das kognitive Interview sollte auf eine Gesamtdauer von 60 bis maximal 90 Minuten angelegt sein, da sonst die Konzentration und Motivation auf Seiten der Befragten (und der Interviewer) deutlich nachlässt. In Abhängigkeit der Anzahl der verwendeten kognitiven Techniken können in diesem Zeitraum gewöhnlich zwischen 20 und 25 Fragen bzw. Items getestet werden. Die Testpersonen im Pretest sollten generell die gleichen Merkmale besitzen wie die Befragungspersonen in der Hauptstudie (bezüglich Alter, Geschlecht, Bildung, etc.). Üblicherweise arbeitet man bei kognitiven Interviews mit Quotenstichproben. Das Ziehen von Zufallsstichproben ist nicht notwendig, da es bei kognitiven Interviews primär darum geht, Fragenprobleme aufzudecken, nicht jedoch eine möglichst genaue Schätzung abzugeben, wie häufig diese Probleme in der Grundgesamtheit auftreten.

5. Wie wertet man kognitive Pretests aus?

Vor der Auswertung von kognitiven Pretests ist es hilfreich, die einzelnen Aufzeichnungen zu transkribieren und eine fallspezifische Auflistung aller Befragtenäußerungen zu erstellen. Eine solche Auflistung beinhaltet folgende Informationen: (1) die Antworten auf die getesteten Fragen, (2) spontane Äußerungen der Testperson zur Frage, (3) die Antworten und Reaktionen der Testperson auf die kognitiven Nachfragen, und (4) Anmerkungen des Testleiters.

Für die Auswertung von kognitiven Interviews stehen verschiedene Verfahren zur Verfügung. Das einfachste und schnellste Verfahren besteht in einer informellen Analyse der Daten, wobei der

Auswerter für jede Äußerung des Befragten entscheidet, ob diese auf ein Frage-Problem hindeutet oder nicht. Dieses Vorgehen birgt allerdings die Gefahr einer sehr subjektiven Einschätzung von Frage-Problemen und einer fehlenden Nachvollziehbarkeit der Pretestergebnisse.

Die formelle Analyse der Daten kann entweder quantitativ oder qualitativ erfolgen. Bei der quantitativen Auswertung von kognitiven Interviews kommen Coding-Schemata zum Einsatz (z.B. DeMaio & Landreth, 2004). Hier werden für das Interviewer- und Befragtenverhalten (bzw. für deren verbale Äußerungen) Codes vergeben wie z.B. (1) Befragter hat Schwierigkeiten mit dem Verständnis der Frage, (2) Befragter versteht bestimmte Wörter nicht, (3) Befragte verstehen die Frage unterschiedlich oder (4) Befragte haben Schwierigkeiten, sich an den Sachverhalt, der in der Frage angesprochen wird, zu erinnern. Bei der quantitativen Auswertung von Interviews handelt es sich also um ein systematisches und objektives Vorgehen, das zum einen die Überprüfung der Interrater-Reliabilität zulässt und zum anderen eine quantitative Datenanalyse ermöglicht (z.B. die Auszählung der Häufigkeit von unterschiedlichen Problemtypen). Die entscheidenden Nachteile dieses Verfahrens bestehen in der Gefahr des Informationsverlusts durch die Reduktion von Textdaten auf Codes und durch die Tatsache, dass Verbesserungsvorschläge nicht anhand der Codes erarbeitet werden können (hierzu muss man zurück zu den verbalen Daten).

Bei der qualitativen Auswertung von kognitiven Interviews bedient man sich qualitativer Methoden, wie z.B. der Constant Comparative Method (CCM). Diese Methode eignet sich besonders zur explorativen Datenanalyse und zur Generierung von Hypothesen. Die CCM umfasst drei Arbeitsschritte:

- 1 Open Coding: Die verbalen Daten der Befragten werden offen nach Themen kodiert und es werden erste Kategorien gebildet.
- 2 Axial Coding: Es wird versucht, die aus den verbalen Daten konstruierten Kategorien zu integrieren und es wird überprüft, ob es Gruppenunterschiede in der Zuweisung der Kategorien gibt.
- 3 Selective Coding: Im letzten Schritt werden übergeordnete Themen spezifiziert, welche die Kategorien verbinden. Dabei wird auch eine Hypothese oder eine Theorie formuliert, welche die Phänomene beschreibt, die eine Survey-Frage erfasst.

Eine gute Darstellung der Anwendung dieser Methode im Rahmen der Auswertung von kognitiven Interviews findet sich bei Ridolfo und Schoua-Glusberg (2011).

6. Literaturverzeichnis

- Behr, D., Braun, M., Kaczmirek, L., & Bandilla W. (2013). Testing the validity of gender ideology items by implementing probing questions in web surveys. *Field Methods*, 25, 124-141.
- Beatty, P. C., & Willis, G. B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly*, 71, 287-311.
- Blair, J., & Conrad, F. G. (2011). Sample size for cognitive interview pretesting. *Public Opinion Quarterly*, 75, 636-658.
- Campanelli, P. (2008). Testing survey questions. In E. de Leeuw, J. Hox, & D. Dillman (Eds.), *International Handbook of Survey Methodology* (pp. 176-200). New York: Erlbaum.
- Converse, J. M., & Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnaire*. Beverly Hills, CA: Sage.

- DeMaio, T., & Landreth, A. (2004). Do different cognitive interview methods produce different results? In S. Presser, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, J. M. Rothgeb, & E. Singer (Eds.), *Methods for testing and evaluating survey questions*. (pp. 88-108). New York: Wiley.
- Faulbaum, F., Prüfer, P., & Rexroth, M. (2009). *Was ist eine gute Frage? Die systematische Evaluation der Fragenqualität*. Wiesbaden: VS Verlag.
- Porst, R. (2000). *Praxis der Umfrageforschung*. Stuttgart: Teubner.
- Porst, R. (2014). *Fragebogen. Ein Arbeitsbuch*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Prüfer, P., & Rexroth, M. (1996). Verfahren zur Evaluation von Survey-Fragen: Ein Überblick. *ZUMA-Nachrichten*, 39, 95-116.
- Prüfer, P., & Rexroth, M. (2005). Kognitive Interviews. *ZUMA How-to-Reihe*, Nr. 15.
- Ridolfo, H., & Schoua-Glusberg, A. (2011). Analyzing cognitive interview data using the constant comparative method of analysis to understand cross-cultural patterns in survey data. *Field Methods*, 23, 420-438.
- Sudman, S., & Bradburn, N. M. (1982). *Asking questions: A practical guide to questionnaire design*. San Francisco, CA: Jossey-Bass.
- Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage.