

Feldexperimente in der empirischen Sozialforschung: Probleme und Lösungen an einem Beispiel aus der Umfragemethodologie

Taddicken, Monika

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

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

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Taddicken, M. (2010). Feldexperimente in der empirischen Sozialforschung: Probleme und Lösungen an einem Beispiel aus der Umfragemethodologie. *Sozialwissenschaftlicher Fachinformationsdienst soFid*, Methoden und Instrumente der Sozialwissenschaften 2010/1, 35-43. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-205064>

Nutzungsbedingungen:

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

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

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

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

Feldexperimente in der empirischen Sozialforschung: Probleme und Lösungen an einem Beispiel aus der Umfragemethodologie

Monika Taddicken

Einleitung

Der vorliegende Beitrag beschäftigt sich mit den Herausforderungen eines Feldexperiments aus der Umfragemethodologie. Zielsetzung des experimentellen Designs war die Evaluierung von Methodeneffekten bei Web-Befragungen mittels eines empirischen Vergleichs alternativer Befragungsmethoden (vgl. Taddicken 2008). Als Methodeneffekte werden Verzerrungen von Antworten und anderen Messergebnissen aufgrund der gewählten Befragungs- bzw. Messmethode bezeichnet. In der Regel handelt es sich dabei um vergleichsweise geringe Verzerrungen (Groves et al. 2004: 162). Da Methodeneffekte aber methodenimmanent, also untrennbar mit einer Erhebungsmethode verbunden sind und damit systematisch, werden sie vereinzelt als bedeutsamste Quelle von Messfehlern beschrieben (vgl. Fiske 1982: 78, Cote & Buckley 1987: 315, Groves 1989: 502). Während bezüglich der Web-Befragung häufig die mit der Auswahl dieser Befragungsmethode auftretenden Probleme der Stichprobenqualität und der erreichbaren Repräsentativität thematisiert wurden, standen die Messfehler, insbesondere die Methodeneffekte zunächst abseits des Fokus.

Dieses Forschungsdesiderat war der Anlass für die Durchführung eines empirischen Methodenvergleichs, bei dem die Web-Befragung mit den Befragungsmethoden Telefon und Postalisch verglichen wurde. Dieser wurde als Feldexperiment mittels eines Test-Retest-Ansatzes realisiert. Im Folgenden werden sowohl die Gründe für diese Entscheidung als auch die damit verbundenen Probleme sowie deren konkrete Lösungen dargelegt. Die Ergebnisse des Methodenvergleichs sind dabei an dieser Stelle nicht von Interesse. Diese sind neben einer ausführlichen theoretischen Behandlung bei Taddicken (2008) zu finden.

Forschungsproblem

Unter Methodeneffekten können ganz allgemein „Auswirkungen der Datenerhebungsmethode auf die Ergebnisse der Messungen“ (Schnell et al. 2005: 160) verstanden werden. Prinzipiell setzt sich nach den Annahmen der klassischen Testtheorie der mittels eines Messinstruments gemessene Wert aus dem sog. ‚wahren Wert‘ und einem spezifischen Messfehler zusammen (vgl. z. B. Lord & Novick 1968, Kranz 1998, Lienert & Raatz 1998). Aus testtheoretischer Perspektive erscheint es somit möglich, aus einem Abgleich des Messergebnisses mit dem ‚wahren Wert‘ auf die Höhe eines Messfehlers zu schließen. Die Sozialforschung sieht sich hier jedoch mit dem Problem konfrontiert, dass die interessierenden Merkmale und damit Gegenstand der Datenerhebungen typischer Weise Meinungen, Einstellungen, Bewertungen und Werthaltungen sind. Es handelt sich also um innere Merkmale, die von außen nicht beobachtet werden können. Allenfalls wäre es möglich, das Verhalten eines Menschen von außen zu erfassen und darüber auf zugrundeliegende innere Merkmale wie

Einstellungen zu schließen. Hier besteht jedoch die Schwierigkeit der komplexen und multi-kausalen Zusammenhänge.

Somit bleibt der Umfragemethodologie zur Evaluierung von Methodeneffekten einer einzelnen Befragungsmethode lediglich der Vergleich verschiedener Methoden. Ein solches Vorgehen entspricht zudem der Alltagsrealität eines empirischen Forschers, der beim Designen seiner Befragungsstudie stets eine Auswahl einer (oder mehrerer) Befragungsmethoden aus dem Kanon der Methoden zu treffen hat.

Forschungsdesign

Um in einem Vergleich mehrerer Methoden die Methodeneffekte bestimmen zu können, bedarf es eines experimentellen Ansatzes, bei dem die Befragungsmethode als unabhängige Variable systematisch variiert wird, um ihren Einfluss auf die Antworten (abhängige Variablen) zu messen. Weitere Erklärungsgrößen müssen dabei ausgeschlossen werden, z. B. indem sie konstant gehalten werden. Insbesondere ist darauf zu achten, dass das Instrument gleich bleibt. Die Auswahl, Reihenfolge und Formulierung sowohl der Fragen als auch der Antwortoptionen darf die Antworten nicht unsystematisch beeinflussen. Aber auch variierende Einflüsse der Befragungsumgebung können als Störfaktoren auftreten. Während ein identischer Fragebogen relativ leicht realisierbar ist, können verschiedenartige Umgebungseinflüsse praktisch nur im Labor unter künstlichen Umweltbedingungen konstant gehalten werden. Wie bei jedem Experiment besteht hier das sog. Validitäts-Dilemma, bei dem die externe und die interne Validität im Hinblick auf das konkrete Forschungsziel gegeneinander abzuwägen sind (vgl. Bortz & Döring 2002: 56f., Bortz 2005: 8). Für die methodologische Umfrageforschung ist das Feldexperiment zu bevorzugen, da die Bedeutsamkeit der voll-biotischen Bedingungen hier als besonders hoch einzustufen ist. Im weitaus überwiegenden Anteil der Befragungen in der empirischen Sozialforschung handelt es sich um Erhebungen im Feld, bei denen sich die Probanden in ihrer natürlichen Umgebung befinden. In anderen Disziplinen, z. B. der Psychologie, sind Labor-Befragungen üblicher.

Allerdings folgt aus der Entscheidung für ein Feldexperiment, dass auftretende Änderungen der abhängigen Variablen, sprich der Antworten, auch Folge von störenden Umwelteinflüssen sein können. Dies ist insbesondere bedeutsam bei Befragungen, in die kein Interviewer direkt involviert ist, wie dies bei der Web-Befragung und der postalischen Befragung der Fall ist. So kann weder die Identität überprüft noch die Anwesenheit Dritter ausgeschlossen werden. (Im Übrigen stellt dies auch besondere Anforderungen an die Verständlichkeit des Fragebogens, weil Rückfragen nicht direkt geklärt werden können, vgl. hierzu Taddicken 2005).

Neben dem Einfluss der systematisch variierten unabhängigen Variable Befragungsmethode und der angesprochenen Störfaktoren, kann der Kausalitätsschluss auch durch eine Änderung der ‚wahren Werte‘ gefährdet sein. Wie gesagt, kann man testtheoretischen Überlegungen folgend annehmen, dass sich Antworten einer Befragung zusammensetzen aus dem ‚wahren Wert‘ und den Messfehlern. Neben den obigen Überlegungen zu den konstant zu haltenden Messfehlerquellen (mit Ausnahme der Methode als UV) und weiteren Störfaktoren der situativen Umgebung ist es von herausragender Bedeutung, dass die ‚wahren Werte‘ als gleichbleibend angenommen werden können. Dies kann realisiert werden über eine mehrmalige Befragung derselben Personen (Test-Retest-Ansatz) (vgl. Groves 1989: 503) oder mittels der Split-Ballot-Technik, bei der mehrere Substichproben mit unter-

schiedlichen Befragungsinstrumenten bzw. -methoden befragt werden (vgl. Schnell et al. 2005: 350f.). Beide Möglichkeiten verfügen über Vor- und Nachteile.

Split-Ballot-Technik versus Test-Retest-Ansatz

Bei der Verwendung der Split-Ballot-Technik, also der Befragung verschiedener Substichproben, ist von entscheidender Bedeutung, dass die ‚wahren Werte‘ in den jeweiligen Substichproben gleich verteilt sind. Ansonsten wäre die Unterschiedlichkeit der Messwerte nicht bzw. nicht nur auf die Methodeneffekte der jeweils eingesetzten Befragungsmethoden zurückzuführen, sondern auch auf die unterschiedlichen ‚wahren Werte‘ in den Stichproben. Dabei ist zu beachten, dass die Gleichverteilung hinsichtlich der interessierenden Variable bestehen muss. Da diese in der Regel nicht bekannt sind, hilft man sich hier – wie auch bei Fragen der Repräsentativität – damit, Hilfsvariablen gleich zu verteilen. Dies sind typischerweise soziodemographische Merkmale (Alter, Geschlecht, Bildung, Einkommen etc.). Für Web-Befragungen ist darauf zu achten, dass ausschließlich Internetnutzer befragt werden, auch in den Stichproben mit alternativen Befragungsmethoden. Obwohl mittlerweile über zwei Drittel der deutschen Bevölkerung ab 14 Jahren online sind und trotz aller Annäherungen der soziodemographischen Kennwerte der Internetnutzer an die deutschen Durchschnitte, sind die Internetnutzer noch immer jünger, männlicher und gebildeter als die deutsche Bevölkerung (vgl. AGOF 2009). Daneben unterscheiden sich die Internetnutzer aber auch in psychographischen Merkmalen von den Nicht-Internetnutzern (vgl. z. B. die Typologie von Gerhards & Mende 2005). Dies ist insbesondere für Merkmale anzunehmen, die in Zusammenhang mit dem Internet stehen, wie z. B. eine allgemeine Technikaffinität oder Innovationsfreude. Allerdings kann dies auch generelle Persönlichkeitsaspekte, wie etwa Offenheit oder Flexibilität betreffen. Daher dürfen bei der Split-Ballot-Technik grundsätzlich nur Personen befragt werden, die zu den Internetnutzern zählen.

Um weiterhin ausschließen zu können, dass sich die Substichproben nicht systematisch bezüglich ihrer Teilnahmebereitschaften unterscheiden, müssten alle Probanden grundsätzlich dazu bereit sein, an allen im Forschungsdesign eingesetzten Befragungsmethoden teilzunehmen. Allerdings kann auch dadurch nicht zuverlässig einem systematischen Nonresponse entgegengewirkt werden. So bleibt fraglich, ob nicht doch die Vorlieben für oder die Abneigung gegen bestimmte Befragungsmethoden für einen Ausfall bestimmter Personengruppen sorgen (bei Web-Befragungen z. B. aufgrund einer allgemeinen Skepsis gegenüber der Sicherheit der Datenübertragung im Internet). So hat eine Meta-Analyse von Grandcolas et al. (2003) gezeigt, dass hohe und ungleich über die Befragungsmethoden hinweg verteilte Verweigerungsquoten systematische Ausfälle vermuten lassen, insbesondere bei der Web-Befragung, sodass dadurch die Parallelität der Stichproben möglicherweise stark eingeschränkt wird (vgl. Grandcolas et al. 2003: 542f.). Auch der Ausfall während der Befragung, also der Befragungsabbruch, kann zu systematischen Differenzen zwischen den Stichproben führen. Durch derartige systematische Effekte kann eine Gleichverteilung der ‚wahren Werte‘ aufgehoben werden.

Insgesamt ist die Split-Ballot-Technik daher für empirische Methodenvergleiche, insbesondere bei Berücksichtigung der Web-Befragung, durch viele Unsicherheiten gekennzeichnet.

Bei der Alternative Test-Retest-Ansatz werden dieselben Befragten mehrmals mit demselben Fragebogen befragt. Dennoch ist dadurch nicht gesichert, dass die ‚wahren Werte‘ zu den verschiedenen Befragungszeitpunkten gleichverteilt sind. So muss zunächst befürchtet werden, dass die erste Be-

fragung das Antwortverhalten für die jeweiligen Fragestellungen in der nachfolgenden Befragung beeinflusst. Die Erinnerung an die gestellten Fragen und an die abgegebenen Antworten kann zu Verzerrungen führen, da Probanden prinzipiell dazu neigen, möglichst konsistent zu antworten (vgl. van Meurs & Saris 1989: 134). Dieses Problem besteht verstärkt bei einem Ansatz zur Evaluierung von Methodeneffekten, da das Instrument vollkommen identisch zu halten ist (anders als bei Instrumententests). Um diesen Erinnerungseffekten vorzubeugen, sollte der zeitliche Abstand zwischen den einzelnen Befragungen möglichst lang sein.

Allerdings darf der Zeitraum zwischen den Wellen auch nicht zu lang gewählt werden, da es zu einer Änderung der ‚wahren Werte‘ kommen kann, wie etwa aufgrund eines bedeutsamen gesellschaftspolitischen Ereignisses zwischen den Befragungswellen. Diese Gefahr ist insbesondere dann gegeben, wenn Merkmale abgefragt werden, die als individuell instabil zu bewerten sind. Das trifft auch auf Meinungen und Einstellungen zu, die die Probanden während der ersten Befragung gebildet haben, die also nicht vorab bestanden. Es können jeweils Lerneffekte entstehen aufgrund der Auseinandersetzung mit diesen Fragen in der ersten Welle, welche die ‚wahren Werte‘ beeinflussen.

Welcher Zeitraum als lang genug angenommen werden kann, um Erinnerungseffekten vorzubeugen, und als kurz genug, um Lerneffekte zu verhindern, hängt davon ab, wie stabil die abgefragten Merkmale sind und wie lang der Erinnerungszeitraum ist. Ob ein Test-Retest-Ansatz verwendet werden kann, hängt somit auch vom Inhalt der Fragestellungen ab (vgl. Groves 1991: 17).

Schließlich verlangt ein Test-Retest-Ansatz den Probanden eine besonders hohe Motivation ab. Durch die Wiederholung der Befragung ist der Aufwand für die Probanden höher als bei einmaligen Befragungen. Mangelnde Teilnahmebereitschaften sind ein typisches Problem bei Test-Retest-Ansätzen. Zudem kommt es häufig zu Befragungsabbrüchen aufgrund des identischen Instruments. Hier muss durch geeignete Maßnahmen versucht werden, die Motivation der Probanden zu erhöhen.

Trotz der auch beim Test-Retest-Ansatz auftretenden Probleme, ist dieser für den angestrebten Methodenvergleich als die bessere Alternative anzusehen. Die Gefahr ungleich verteilter ‚wahrer‘ Werte in verschiedenen Substichproben wurde aufgrund der oben genannten Gründe als höher eingestuft als die potenziellen Beeinträchtigungen der Antworten, die durch einen Test-Retest-Ansatz entstehen (vgl. Taddicken 2008: 165).

Lösungswege

Die Umsetzung des Test-Retest-Ansatzes in ein Forschungsdesign muss also die beschriebenen Herausforderungen meistern. Am Beispiel eines empirischen Methodenvergleichs mit den Methoden Web, Telefon und Post werden folgend konkrete Lösungsmöglichkeiten aufgezeigt.

Zunächst wurde entschieden, die mehrmalige Befragung der Probanden auf zwei Wellen zu begrenzen. Die Probanden wurden also jeweils nur mit zwei verschiedenen Methoden befragt, obwohl drei Methoden verglichen wurden (vgl. Taddicken 2008: 167). Dies ist der Einschätzung geschuldet, dass die dreimalige Beantwortung desselben Fragebogens Erinnerungseffekte in nicht mehr zu korrigierenden Ausmaße bedingen würde (vgl. dazu van Meurs & Saris 1989: 144ff.). Insofern waren zwei Experimentalgruppen zu bilden: eine Gruppe, in der die Probanden einmal mit einer Web-Befragung und einmal telefonisch befragt werden, und eine Gruppe, in der einmal ein Web- und einmal ein schriftlicher Fragebogen zu beantworten ist.

Um weiterhin Reihenfolgeeffekte zu kontrollieren, wurde innerhalb der Experimentalgruppe die Reihenfolge variiert. Dazu wurden die beiden Experimentalgruppen in zwei Untergruppen (a und b) gesplittet. Eine Untergruppe (Stichproben 1a und 2a) konnte somit jeweils in der ersten Welle per

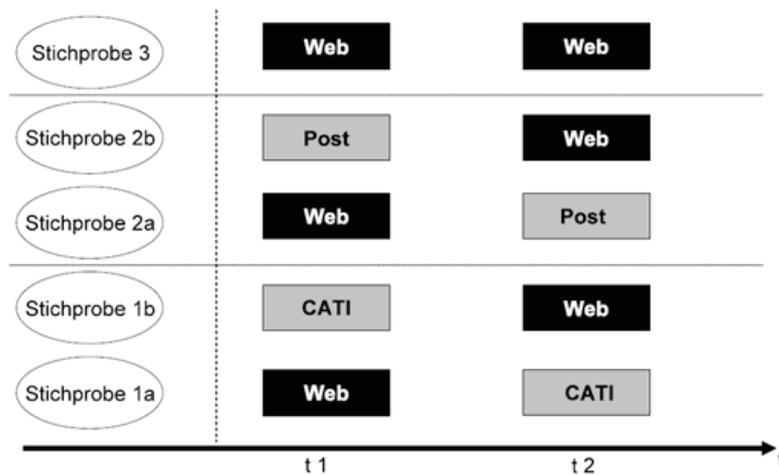


Abbildung 1: Darstellung des entwickelten Forschungsdesigns (Quelle: Taddicken 2008: 168).

Web befragt werden, während die andere (Stichproben 1b und 2b) jeweils in der zweiten Welle die Web-Befragung erhielt. Außerdem wurde eine Kontrollgruppe eingesetzt, durch die weitere mögliche zeitliche Effekte aufgedeckt werden können. Die Kontrollgruppe (Stichprobe 3) wurde zweimal mittels Web befragt (vgl. Abbildung 1).

Inhaltlich wurde darauf geachtet, ausschließlich vergleichsweise stabile Variablen abzufragen (konkret waren dies Werthaltungen, Bindungen an soziale Normen sowie alltägliche Verhaltensweisen). Außerdem wurden vorrangig etablierte Skalen mit dokumentierten hohen Reliabilitäten verwendet. Leider werden in Skalendokumentationen häufig nicht die eigentlich interessierenden Test-Retest-Reliabilitäten berichtet, so dass oftmals auf die interne Konsistenz als Maß ausgewichen werden muss.

Bezüglich des zeitlichen Abstands zwischen den beiden Befragungen wurde zunächst Empfehlungen aus der Literatur gefolgt. So schlagen Saris (1989: 165) bzw. van Meurs & Saris (1989: 146) vor, drei bis vier Wochen Zeitabstand einzuhalten. Dieser Zeitraum wurde allerdings nach Durchführung eines Pretests um eine Woche auf vier bis fünf Wochen verlängert. Explorative Nachbefragungen der Pretestteilnehmer hatten gezeigt, dass sich diese noch an die Fragen und gegebene Antworten der ersten Welle erinnerten (vgl. Taddicken 2008: 168f.).

Um dem Problem mangelnder Motivation und Antwortbereitschaft aufgrund des mit der Studie für die Probanden verbundenen Aufwands entgegenzuwirken, wurde zum einen deutlich auf den universitären Kontext des Projekts hingewiesen und über eine Vorstellung des Forschers an eine altruistisch-motivierte Teilnahme appelliert, zum anderen wurde sowohl über ein Gewinnspiel als auch über Gutscheine für alle ‚Finisher‘ incentiviert.

Rekrutiert wurden die Probanden aus pragmatischen Gründen über ein Online-Access-Panel. Eine Internetnutzer-Repräsentativität konnte aus diesem Grund nicht angestrebt werden. Die Panel-Mitglieder wurden über eine e-Mail auf das Projekt aufmerksam gemacht. Über einen Link konnten sie auf eine Informationsseite gelangen, die das Projekt vorstellte (jedoch nicht seine methodische Zielsetzung). Hier war es den interessierten Mitgliedern auch möglich, ihre Kontaktdaten für eine Teilnahme zu hinterlassen. Da die Zuteilung auf die Experimentalgruppen zufällig erfolgen sollte, war es notwendig, dass neben Namen und e-Mail-Adresse auch Post-Adresse und Telefonnummer angegeben wurden.

Erfahrungen

Wie gesagt wurde als zeitlicher Abstand zwischen den Befragungswellen ein Zeitraum von vier bis fünf Wochen gewählt. Dieser wurde als lang genug eingeschätzt, um Erinnerungseffekten entgegen zu wirken, und als kurz genug, um ‚normale‘ zeitlich bedingte Änderungen der als relativ stabil anzusehenden abgefragten Merkmale auszuschließen. Zwar ist im Einzelfall nur schwer überprüfbar, ob individuelle Erinnerungseffekte aufgetreten sind. Mit Hilfe der Kontrollgruppe aber wurde getestet, ob insgesamt systematische Verschiebungen in Abhängigkeit der Zeit aufgetreten sind. Obgleich sich einige Auffälligkeiten gezeigt haben, war insgesamt nicht davon auszugehen. Dass die Zeit als Störfaktor trotzdem nicht ausgeschaltet werden konnte, zeigte sich im durchgeführten Methodenvergleich anhand der Abfrage individueller reflexiver Werthaltungen nach Klages (1992) und Hermann (2003). Obwohl gerade Werthaltungen, die als relativ konstante präskriptive Überzeugungen von Individuen definiert werden können, als stabile Konstrukte anzusehen sind, waren die Test-Retest-Reliabilitäten vergleichsweise niedrig (Pearsons Korrelation $r < .5$). Die Antworten auf diese Skala zeichneten sich also durch eine nur geringe Stabilität im Zeitverlauf aus. Dies kann zum einen auf eine mögliche schlechte Güte des Messinstruments zurückgeführt werden, zum anderen aber muss eine Änderung der ‚wahren Werte‘ befürchtet werden. Die innere Auseinandersetzung mit den eigenen Werthaltungen, angestoßen durch die Befragung in Welle 1, kann zu veränderten und/oder bewussteren Einstellungen geführt haben.

Ferner wurde ein Einfluss mangelnder Motivation der Probanden befürchtet. Der Feldverlauf bestätigte diese Befürchtung jedoch nicht. Zwar war die Bereitschaft, sich für das aufwändige Befragungsprojekt zu registrieren gering; von 5.000 angeschriebenen Panel-Mitgliedern registrierten sich nur 816 Personen. Allerdings ist darauf hinzuweisen, dass die Studie zum einen im Sommer durchgeführt wurde, was erfahrungsgemäß zu einer schlechteren Erreichbarkeit führt, zum anderen hat vermutlich auch die umfangreiche Abfrage der Kontaktdaten (inklusive Adresse, Telefonnummer) abschreckend gewirkt. Das Durchhaltevermögen der vollständig registrierten Personen ($n=799$) aber war sehr hoch. So betrug die durchschnittliche Rücklaufquote über alle Befragungsmethoden hinweg 73,2%.

Weiterhin war zu befürchten, dass die Bekanntheit der Fragen in der zweiten Welle dazu führt, die Fragen weniger konzentriert zu beantworten. Dies gilt vor allem, wenn in der zweiten Welle die Web-Befragung oder die postalische Befragung stattgefunden hat, da kein Interviewer motivierend auf den Probanden einwirken konnte. Es wurde daher kontrolliert, ob Fragen einfach ‚durchgeklickt‘ bzw. ‚durchgekreuzt‘ wurden, ob also ganze Itembatterien dieselben Antwortwerte aufweisen. Dies war nicht der Fall. Auch weisen die Abbrecherquoten in den jeweiligen Befragungen nicht auf einen Motivationsmangel der Probanden hin. So waren die Abbrecherquoten in der Web-Befragung in der

zweiten Welle mit 1,4 % sogar geringer als noch in der ersten Welle (3,4 %). Am Telefon konnte in beiden Wellen eine Abbrecherquote von 0 % erreicht werden. Für die postalische Befragung war eine Feststellung der Abbrecherquoten nicht möglich.

Schließlich war ein wesentlicher Störfaktor in dem durchgeführten Methodenvergleich die fehlende Kontrolle der Befragungssituation. So wurden die Befragungssituationen weder kontrolliert noch waren sie einsehbar. Die Vergleichbarkeit der Befragungssituation, die unter Umständen Einfluss auf das Antwortverhalten haben kann, konnte insofern nicht gewährleistet werden. Dies schlägt sich insbesondere störend nieder, weil unterschiedliche Befragungsmethoden unterschiedliche Befragungssituationen bedingen. Bei der telefonischen Befragung wurden die meisten Probanden zu Hause kontaktiert. Nur wenige Teilnehmer wurden über ein Mobiltelefon angerufen. Auch bei der postalischen Befragung ist zu vermuten, dass diese überwiegend im häuslichen Umfeld beantwortet wurde. Allerdings könnte die Beantwortung auch außer Haus vorgenommen worden sein, z. B. bei der Arbeit oder in öffentlichen Verkehrsmitteln. Es ist weiter unklar, in welcher Umgebung der Web-Fragebogen ausgefüllt wurde. Dies kann zu Hause, aber auch in einem Internetcafé, bei der Arbeit bzw. in der Schule oder Uni geschehen sein. Durch die unkontrollierten Erhebungssituationen kann ein Einfluss der Umwelt während der Befragung nicht ausgeschlossen werden. So kann es zu Ablenkungen des Probanden gekommen sein (z. B. durch einen Fernseher, Türklingeln, Kinder oder Haustiere). Es können sogar weitere Personen anwesend gewesen sein und die Antworten der Probanden (bewusst oder unbewusst) beeinflusst haben. Inwiefern sich dadurch insgesamt Beeinflussungen der Antworten ergeben haben, kann insgesamt nicht beurteilt werden. Eine in den Fragebogen integrierte Frage nach der konkreten Situation und Umgebung der Probanden könnte hier weitere Erkenntnisse liefern. Ein Ausschluss dieses Störfaktors ist praktisch nur bei Durchführung des Methodenvergleichs im Labor möglich.

Fazit

Der Fokus dieses Beitrags lag auf der Darstellung der Herausforderungen des Methodenvergleichs als Feldexperiment und deren Begegnung. Für die inhaltlichen Ergebnisse, die tatsächlich Methodeneffekte der Web-Befragung bestätigen, sei auf Taddicken (2008) verwiesen.

Für eine Evaluierung von Methodeneffekten ist – wie dargelegt – ein Vergleich mehrerer Methoden notwendig, da in der Regel kein ‚wahrer Wert‘ bekannt ist, der vom erhobenen Wert abgezogen werden könnte. Solche Methodenvergleiche stellen prinzipiell experimentelle Ansätze dar. Die abhängigen Variablen, in diesem Fall die Antworten der Probanden, werden erhoben und in Abhängigkeit von den unabhängigen Variablen (hier: die Befragungsmethoden), die systematisch variiert werden, miteinander verglichen. Wenn sämtliche Störfaktoren ausgeschlossen werden können, sind die Differenzen in den Antworten, in den abhängigen Variablen, auf die unterschiedlichen Befragungsmethoden zurückzuführen. Es ist jedoch in der Praxis nicht möglich, sämtliche Störfaktoren zu beseitigen.

In diesem Zusammenhang wurden als die größten Probleme zeitliche Effekte (Erinnerungs- und Lerneffekte) und eine fehlende Teilnahmemotivation als Folge des Test-Retest-Ansatzes sowie die mangelnden Kontrollmöglichkeiten aufgrund der Realisierung als Feldexperiment festgehalten. Es wurden Wege aufgezeigt, diesen Problemen zu begegnen und Erfahrungswerte hinsichtlich dieser Lösungen berichtet.

Dabei hat sich gezeigt, dass die zeitliche Instabilität bestimmter Merkmale das größte Problem darstellte. Diese ist vermutlich auf Lerneffekte durch die Auseinandersetzung mit den Fragen in der ersten Welle zurückzuführen. Erinnerungs- und Motivationsprobleme konnten dagegen nicht nachgewiesen werden. Wie erwähnt, bedeutet dies nicht, dass diese im Einzelfall nicht aufgetreten sind. Es lässt aber hoffen, dass man den vielen Herausforderungen eines experimentellen Test-Retest-Ansatzes im Feld mit Hilfe eines darauf abgestimmten Forschungsdesigns gerecht werden kann.

Literatur

- AGOF (Arbeitsgemeinschaft Online-Forschung) (2009): *Berichtsband zur internet facts 2009-III*. Frankfurt. <http://www.agof.de> (18.12.2009).
- Bortz, Jürgen (2005): *Statistik für Human- und Sozialwissenschaftler*. Heidelberg, 6. Aufl.
- Bortz, Jürgen & Döring, Nicola (2002): *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Berlin, Heidelberg, 3. Aufl.
- Cote, Joseph A. & Buckley, M. Ronald (1987): Estimating Trait, Method, and Error Variance: Generalizing across 70 Construct Validation Studies. In: *Journal of Marketing Research*, 24 (3), S. 315-318.
- Fiske, Donald W. (1982): Convergent – Discriminant Validation in Measurements and Research Strategies. In: Brinberg, David & Kidder, Louise, H. (Hrsg.): *Forms of Validity in Research*. San Francisco, S. 77-93.
- Gerhards, Maria & Mende, Annette (2005): Offliner: Zwischen interessierter Annäherung und bewusster Distanz zum Internet. In: *Media Perspektiven*, 8, S. 380-395.
- Groves, Robert M. (1989): *Survey Errors and Survey Costs*. New York.
- Groves, Robert M. (1991): Measurement Errors Across the Disciplines. In: Biemer, Paul P.; Groves, Robert M.; Lyberg, Lars E.; Mathiowetz, Nancy A. & Sudman, Seymour (Hrsg.): *Measurement Error in Surveys*. New York, S. 1-25.
- Groves, Robert M.; Fowler, Floyd J., Jr.; Couper, Mick P.; Lepkowski, James M.; Singer, Eleanor & Tourangeau, Roger (2004): *Survey Methodology*. Hoboken.
- Hermann, Dieter (2003): *Werte und Kriminalität. Konzeption einer allgemeinen Kriminalitätstheorie*. Wiesbaden .
- Klages, Helmut (1992): Die gegenwärtige Situation der Wert- und Wertwandelforschung. Probleme und Perspektiven. In: Klages, Helmut; Hippler, Hans-Jürgen & Herbert, Willi (Hrsg.): *Werte und Wandel. Ergebnisse und Methoden einer Forschungstradition*. Frankfurt am Main, S. 5-39.
- Kranz, Helger T. (1997): *Einführung in die klassische Testtheorie*. Frankfurt am Main, 4. Aufl.
- Lienert, Gustav A. & Raatz, Ulrich (1998): *Testaufbau und Testanalyse*. Weinheim, 6. Aufl.
- Lord, Frederic M. & Novick, Melvin R. (1968): *Statistical Theories of Mental Test Scores*. Reading, Massachusetts.
- Saris, Willem E. (1989): The Choice of a Research Design for MTMM Studies. In: Saris, Willem E. & van Meurs, Alex (Hrsg.): *Evaluation of Measurement Instruments by Meta-Analysis of Multi-trait Multimethod Studies*. Amsterdam, S. 160-167.
- Schnell, Rainer; Hill, Paul B. & Esser, Elke (2005): *Methoden der empirischen Sozialforschung*. München, Wien, 7. Aufl.

- Taddicken, Monika (2005): Perspektive Deutschland: Mehr als ein gesellschaftspolitisches Meinungsbild?, in: *planung & analyse*, 1, S. 29-34.
- Taddicken, Monika (2008): Methodeneffekte bei Web-Befragung. Einschränkungen der Datengüte durch ein ‚reduziertes Kommunikationsmedium‘?. Reihe: Neue Schriften zur Online-Forschung, Band 5. Köln.
- van Meurs, Alex & Saris, Willem E. (1989): Memory Effects in MTMM Studies. In: Saris, Willem E. & van Meurs, Alex (Hrsg.): *Evaluation of Measurement Instruments by Meta-Analysis of Multi-trait Multimethod Studies*. Amsterdam, S. 134-146.

Zur Person

Dr. Monika Taddicken
Institut für Journalistik und Kommunikationswissenschaft
Allende-Platz 1
20146 Hamburg
E-Mail: monika.taddicken@uni-hamburg.de

Arbeitsschwerpunkte:

Online-Forschung, Umfragemethodologie, Social Web, Medienwirkungs- und Rezeptionsforschung.