

SDM Survey Guidelines

**Antworttendenzen in standardisierten
Umfragen**

Kathrin Bogner & Uta Landrock

Zusammenfassung

Antworttendenzen liegen vor, wenn in standardisierten Umfragen die erhobenen bzw. berichteten Werte der Befragten von den tatsächlichen, wahren Werten systematisch abweichen. Somit sind Antworttendenzen potentielle Fehlerquellen in sozialwissenschaftlichen Umfragen. Antworttendenzen treten auf, wenn Befragte in Interviews nicht nur auf den Inhalt der Frage, den Fragestimulus, reagieren, sondern in ihrer Antwort auch von anderen Faktoren beeinflusst werden. Diese Faktoren finden sich in den verschiedensten Elementen einer Befragung, sie können in der Person des Befragten, in der Interviewsituation sowie im Erhebungsinstrument liegen.

Antworttendenzen, die in der Person des Befragten liegen, sind oft Folge von Satisficing: Wenn Befragte bei der Beantwortung von Fragen zum Satisficing neigen, wollen sie ihren kognitiven Aufwand reduzieren; in der Folge können Antworttendenzen wie Akquieszenz/Zustimmungstendenz oder die Tendenz zur Mitte entstehen. Bei der sozialen Erwünschtheit orientieren sich Befragte an sozialen Normen, um so zu antworten, wie sie annehmen, dass es den Erwartungen des Interviewers entspricht. Bei der Tendenz zur Milde/Härte neigen Befragte dazu, unabhängig vom Inhalt der Frage die Extremkategorien in Ratingskalen anzugeben. Ein zweiter Faktor ist die Interviewsituation. Antworttendenzen treten auch dann auf, wenn das Antwortverhalten von der Situation abhängig ist: Es handelt sich um Interviewereffekte, wenn Befragte auf Merkmale oder Verhaltensweisen des Interviewers reagieren. Von Anwesenheitseffekten (Anwesenheit dritter Personen) spricht man dagegen, wenn das Antwortverhalten des Befragten von anwesenden Dritten abhängig ist. Der dritte Einflussbereich für Antworttendenzen ist das Erhebungsinstrument: Die Formulierung der Fragen im Fragebogen, ihre Reihenfolge sowie die Reihenfolge der Antwortalternativen im Fragebogen können z.B. als Ausstrahlungseffekte oder Primacy- bzw. Recency-Effekte das Antwortverhalten beeinflussen und zu Antworttendenzen führen.

Zitierung

Bogner, Kathrin und Landrock, Uta (2015): Antworttendenzen in standardisierten Umfragen. Mannheim, GESIS – Leibniz Institut für Sozialwissenschaften (SDM Survey Guidelines). DOI: 10.15465/sdm-sg_016

Inhaltsverzeichnis

1. Antworttendenzen in standardisierten Umfragen	1
2. Typen von Antworttendenzen	2
2.1 Soziale Erwünschtheit	2
2.2 Akquieszenz	3
2.3 Tendenz zur Mitte und zur Milde/Härte	4
3. Einfluss des Interviewers und der Befragungssituation	5
3.1 Interviewereffekte	5
3.2 Anwesenheit dritter Personen	6
4. Einfluss des Fragebogendesigns	7
4.1 Effekte der Fragereihenfolge	7
4.2 Reihenfolge der Antwortvorgaben	8
Literaturverzeichnis	10

1. Antworttendenzen in standardisierten Umfragen

Unter Antworttendenz, auch als Antwortverzerrung oder im Englischen als response set oder response bias bezeichnet, wird eine systematische Abweichung der berichteten von den tatsächlichen Werten der Befragten verstanden. Somit stellt sie eine potentielle Fehlerquelle in standardisierten sozialwissenschaftlichen Umfragen dar. Die Ursachen für diese Abweichung sind in verschiedenen Elementen der Befragung zu finden: Bestimmte Antworttendenzen, wie beispielsweise soziale Erwünschtheit oder Akquieszenz, können Folge eines spezifischen Antwortverhaltens des Befragten sein. Bei der Akquieszenz etwa kann es sich einerseits um strategisches Antwortverhalten im Sinne eines Satisficing (s.u.) oder andererseits um die Ausprägung einer Persönlichkeitseigenschaft handeln (Krosnick, 1999b). Des Weiteren tragen auch der Frageinhalt und die Gestaltung des Fragebogens sowie Merkmale der Interviewsituation, z. B. die Anwesenheit eines Interviewers oder dritter Personen, zum Auftreten von Antworttendenzen bei. Insgesamt betrachtet, sind Antworttendenzen eine Folge davon, dass Befragungen soziale Prozesse darstellen, in denen Befragte nicht nur auf den Fragestimulus reagieren, sondern auch auf andere Aspekte der Situation (Kahn & Cannell, 1957; Groves et al., 2009).

Der optimale kognitive Prozess, den Befragte bei der Beantwortung einer Frage durchlaufen, umfasst vier Schritte

- (1) Verstehen der Frage,
- (2) Gewinnung der relevanten Information aus dem Gedächtnis,
- (3) Beurteilung der gewonnenen Information hinsichtlich der Vollständigkeit und Relevanz,
- (4) Antwortgabe (Cannell, Miller, & Oksenberg, 1981; Tourangeau, Rips, & Rasinski, 2000)

Jeder einzelne Schritt kann mit hohem kognitivem Aufwand verbunden sein, was in der Summe aller Fragen einer Erhebung unter Umständen eine hohe kognitive Belastung für den Befragten bedeutet (vgl. u.a. Krosnick & Fabrigar, 1997). Dieser so genannten „high road“ der Informationsverarbeitung, Optimizing genannt, steht eine „low road“ - ein an einer Reduzierung von Belastung ausgerichteter Prozess - gegenüber, welcher als Satisficing bezeichnet wird. Während Befragte bei einem schwachen Satisficing noch alle vier kognitiven Schritte durchlaufen, wird bei stärkerem Satisficing der Informationsverarbeitungsprozess verkürzt bzw. bei sehr starkem Satisficing überhaupt nicht vollzogen, beispielsweise wird die Frage gar nicht gelesen. Befragte, die ein Satisficing-Verhalten zeigen, beurteilen die Fragestellung lediglich oberflächlich und berichten die Antwort, die ihnen zuerst in den Sinn kommt oder die ihnen als plausibel erscheint (Krosnick, Nayaran, & Smith, 1996). In der Folge können Antworttendenzen entstehen, beispielsweise wenn Befragte unabhängig vom Frageinhalt die Mittelkategorie wählen, um den kognitiven Prozess zu vereinfachen.

Drei Faktoren bestimmen die Wahrscheinlichkeit, dass ein Befragter bei der Fragebeantwortung eine Satisficing-Strategie wählt: (1) die Schwierigkeit der Aufgabe, (2) die kognitive Fähigkeit und (3) die Motivation des Befragten die Aufgabe zu erfüllen (vgl. u.a. Krosnick et al., 1996; Krosnick, 1999b).

In diesem Beitrag werden die bekanntesten Typen von Antworttendenzen in standardisierten sozialwissenschaftlichen Umfragen behandelt. Diese sind soziale Erwünschtheit (Kapitel 2.1), Akquieszenz (Kapitel 2.2) sowie die Tendenz zur Mitte und zur Milde bzw. Härte (Kapitel 2.3). Des Weiteren wird der Einfluss der Befragungssituation vorgestellt (Kapitel 3.1: Interviewereffekte, Kapitel 3.2: Anwesenheit dritter Personen). Auch die Gestaltung des Fragebogens, das heißt, die Reihenfolge der Fragen (Kapitel 4.1) und die Reihenfolge der Antwortmöglichkeiten (Kapitel 4.2), kann das Auftreten von Antworttendenzen verursachen.

2. Typen von Antworttendenzen

In standardisierten, sozialwissenschaftlichen Umfragen sind die bekanntesten Antworttendenzen soziale Erwünschtheit, Zustimmungstendenz (Akquieszenz), die Tendenz zur Mitte sowie die Tendenz zur Milde bzw. Härte. Auf diese wird im Folgenden näher eingegangen.

2.1 Soziale Erwünschtheit

Soziale Erwünschtheit beschreibt die Tendenz von Befragungsteilnehmern, in einer Befragung eine überwiegend positive Beschreibung der eigenen Person abzugeben (Paulhus, 2002) bzw., allgemeiner gefasst, dem zu entsprechen, was der Interviewer oder andere beteiligte Personen vermeintlich von einem erwarten. Dies kann mittels einer übertriebenen Nennung von erwünschtem Verhalten oder mittels einer untertriebenen Nennung von unerwünschtem Verhalten geschehen. Orientierung bieten dabei soziale Normen (Kreuter, Presser, & Tourangeau, 2008). Ein Beispiel aus vielen westlichen Ländern ist die soziale Norm der Teilnahme an Wahlen, welche als Bürgerpflicht betrachtet wird. Eine Nichtteilnahme wird als Verletzung dieser Norm gewertet, weshalb einige Befragte fälschlicherweise eine Teilnahme berichten, wie ein Vergleich mit der amtlichen Wahlbeteiligung zeigt (Belli, Traugott, & Beckmann, 2001; Holbrook & Krosnick, 2010).

Vor allem bei Fragen zu Themen, die von den befragten Personen als sensibel gewertet werden (z.B. zu Sexualverhalten, Abtreibung oder Drogenmissbrauch), kann die Tendenz einer befragten Person zu sozial erwünschtem Antwortverhalten zu Tage treten (Kreuter et al., 2008). Auch der Modus der Datenerhebung beeinflusst das Auftreten von sozialer Erwünschtheit. Befragte zeigen eher die Tendenz zu sozial erwünschtem Antwortverhalten, wenn die soziale Distanz – hier im Sinne von physischer Nähe bzw. Anwesenheit – zwischen ihnen und dem Interviewer gering ist, wie zum Beispiel in einem persönlichen Interview. Hier besteht für die befragten Personen die Gefahr der Bloßstellung vor dem Interviewer. Liegt jedoch eine große soziale Distanz zwischen Befragtem und Interviewer oder Forscher vor, wie zum Beispiel in einem selbstadministrierten Web Survey, fällt die Tendenz sich positiver darzustellen geringer aus (Tourangeau et al., 2000; Tourangeau & Yan, 2007; Kreuter et al., 2008; Holbrook & Krosnick, 2010; Preisendörfer & Wolter, 2014). Es gibt aber auch Unterschiede zwischen den Modi mit Interviewer. Holbrook, Green und Krosnick (2003) zeigen, dass Befragte im Telefoninterview eher ein Antwortverhalten zeigen, das an sozialen Normen ausgerichtet ist als im persönlichen Interview. Dies führen sie darauf zurück, dass das Vertrauensverhältnis zwischen Interviewer und Befragtem sich am besten in einem persönlichen Interview aufbauen kann, weshalb Befragte in diesem Modus weniger ihre Antworten an sozialen Normen anpassen als in einem Telefoninterview (Holbrook et al., 2003). In einer Validierungsstudie zeigen Preisendörfer und Wolter (2014), dass höher gebildete Befragte, Frauen und Ältere weniger oft unerwünschtes Verhalten zugeben als jüngere Befragte, Männer und niedriger Gebildete. Einen umfassenden Überblick über die Determinanten des sozial erwünschten Antwortverhaltens in Umfragen gibt Krumpal (2013).

Durch sozial erwünschtes Antwortverhalten entstehen systematische Messfehler, welche die Interpretierbarkeit von Randsummen, individuellen Unterschieden und statistischen Zusammenhängen auf Basis dieser Daten gefährden. Daher beschäftigen sich zahlreiche Studien mit der Entwicklung von Skalen zur Messung der Sozialen Erwünschtheit, mit dem Ziel, diese zur Kontrolle der Konfundierung der Daten einzusetzen (für eine Übersicht solcher Arbeiten siehe z.B. Paulhus, 2002; Kemper, Beierlein, Bensch, Kovaleva, & Rammstedt, 2012). Viele der entwickelten Skalen sind allerdings zu lang, um sie standardmäßig in sozialwissenschaftlichen Umfragen zu erheben. Eine ökonomische Messung bietet

eine aus sechs Items bestehende Kurzsкала für soziale Erwünschtheit (KSE-G), die sowohl in deutscher als auch in englischer Sprache entwickelt wurde (Kemper et al., 2012).¹

Durch den Einsatz bestimmter verdeckter Antworttechniken bei sensiblen Fragen ist es möglich, bereits bei der Datenerhebung dem Auftreten von sozial erwünschtem Antwortverhalten entgegenzuwirken. Techniken wie beispielsweise die Randomized-Response-Technique oder die Item-Count-Technik bieten den Befragten die Möglichkeit vertrauliche und anonyme Antworten zu geben. Einige Studien zeigen, dass durch den Einsatz von verdeckten Antworttechniken höhere Raten an unerwünschtem Verhalten erhoben werden als bei einer direkten Befragung. Andere Studien hingegen können die Wirksamkeit nicht belegen. Ein zentrales Problem bei der Verwendung von verdeckten Antworttechniken ist, diese den Befragten so zu erklären, dass sie die Technik verstehen und als plausibel empfinden und ihre wahren Antworten zu den sensiblen Themen geben (vgl. für eine Darstellung der Techniken sowie eine Übersicht über die Studienlage z.B. Krumpal, 2013).

2.2 Akquieszenz

Die Akquieszenz bzw. Zustimmungstendenz gehört zu den Antworttendenzen, die unabhängig vom Inhalt der Frage sind (Krosnick, 1991). Die Befragungspersonen stimmen dabei Aussagen unabhängig vom Inhalt zu, sie geben also die Kategorien „Stimme zu“, „Wahr“ oder „Ja“ an (Krosnick, 1999a). In der Folge können systematische Verzerrungen entstehen (Billiet & Davidov, 2008).

Die Quellen von Akquieszenz können im Stimulus, also in der formalen Gestaltung der Frage selbst oder beim Befragten liegen. Im ersten Fall hilft es, die Fragen umzuformulieren oder anders zu stellen, um Akquieszenz zu vermeiden (Billiet & Davidov, 2008): In einigen Studien wurde festgestellt, dass widersprüchlich formulierte Items (McBride & Moran, 1967) bzw. Items, die aus vielen Worten bestehen (Trott & Jackson, 1967), mit erhöhter Akquieszenz einhergehen können. Es kann also hilfreich sein, solche Items umzuformulieren. Letztendlich können Verzerrungen durch Akquieszenz nur verhindert werden, wenn Zustimmungsskalen wie „stimme zu – stimme nicht zu“, „wahr – falsch“ oder „ja – nein“ vermieden werden (Krosnick, 1999a).

Wenn man annimmt, dass die Ursache für Akquieszenz in der Befragungsperson liegt, finden sich verschiedene Erklärungen. Akquieszenz kann zunächst der Höflichkeit der Befragungsperson geschuldet sein, da es angenehmer ist, zuzustimmen als zu widersprechen (Leech, 1983, zit. n. Krosnick, 1999a). Eine darüber hinausgehende Betrachtungsweise ist, dass die Neigung zu Akquieszenz ein stabiles Persönlichkeitsmerkmal bzw. eine stabile Verhaltenstendenz ist (Krosnick, 1991). Krosnick (1991) zeigt weiterhin, dass Akquieszenz auch die Folge einer Antwortstrategie sein kann, die sich durch Satisficing auszeichnet: Dabei neigen Befragte dazu, nur darüber nachzudenken, ob eine Aussage wahr ist und nicht, ob sie auch falsch sein könnte. Sie überlegen sich also nur Proargumente und keine Kontraargumente. Diese Entscheidungsregel begünstigt Akquieszenz. Anders formuliert: Menschen, die den Aufwand betreiben und sich Gründe überlegen, warum eine Aussage nicht wahr sein sollte oder warum sie ihr nicht zustimmen sollten, neigen weniger zu Akquieszenz (Krosnick, 1991).

Mit Hilfe „ausgeglichener/ balancierter Skalen“ (balanced scales, Billiet & Davidov, 2008) kann Akquieszenz identifiziert und auch kontrolliert werden. Dabei handelt es sich um Skalen, in denen die Hälfte der Items in positiver Richtung und die andere Hälfte in negativer Richtung konstruiert sind, es werden also jeweils eine Aussage und ihr Gegenteil formuliert, so dass Paare sich gegenseitig ausschließender Aussagen entstehen. Diesem Vorgehen liegt die Annahme zugrunde, dass beim Vorliegen von Akquieszenz die Befragten sowohl einer Aussage als auch ihrem Gegenteil zustimmen werden, so dass sich die Antworten ausgleichen – der Mittelwert der Skala wird in solchen Fällen nicht

¹ Siehe auch <http://www.gesis.org/kurzskalen-psychologischer-merkmale/kurzskalen/soziale-erwuenschtheit/>

verzerrt (Krosnick, 1999a; Billiet & Davidov, 2008)². Das ist allerdings nur dann der Fall, wenn man annimmt, dass positiv und negativ formulierte Items gleichermaßen anfällig für Akquieszenz sind. Das scheint aber nicht immer gegeben zu sein (McClendon, 1992, zit. n. Billiet & Davidov, 2008). Der Einsatz ausgeglichener Skalen funktioniert außerdem nur, wenn wirklich gegenteilige Aussagen gefunden werden und der Befragte bei allen Items mit Akquieszenz reagiert (Krosnick, 1999a).

Durch Akquieszenz verzerrte Daten können mit verschiedenen Mitteln korrigiert werden (Krosnick, 1999a): Eine Möglichkeit besteht darin, dass Items in den Fragebogen aufgenommen werden, mit denen man eine in der Persönlichkeit des Befragten liegende Zustimmungstendenz identifizieren kann. Diese Methode beruht auf der Annahme, dass akquieszentes Antwortverhalten ausschließlich auf die Ausprägung eines Persönlichkeitsmerkmals zurückzuführen ist. Das ist jedoch nicht immer gegeben, wie anhand des strategischen Antwortverhaltens, des Satisficing, gezeigt wurde. In solchen Fällen hilft der Einsatz ausgeglichener Skalen, der, wie beschrieben, dazu führt, dass die Mittelwerte von Akquieszenz bereinigt und somit nicht verzerrt werden.

Wer ist akquieszent? Empirische Studien geben Hinweise darauf, dass Befragte mit geringerem sozialem Status und mit geringerer Intelligenz bzw. kognitiven Fähigkeiten besonders zu Akquieszenz neigen (Krosnick, 1999a). Es kann weiterhin gezeigt werden, dass Akquieszenz unter Befragten mit geringerer formaler Bildung höher ist (Rammstedt, Goldberg, & Borg; 2010). Die Neigung zu Akquieszenz wird, ebenso wie andere Ausprägungen des Satisficings, durch die Schwierigkeit der Items, die Fähigkeit des Befragten und durch seine Motivation beeinflusst (Krosnick, 1991).

Für das Auftreten von Akquieszenz in Telefon- und face-to-face-Interviews spielt der Interviewer eine wichtige Rolle (Olson & Bilgen, 2011): Es zeigt sich, dass erfahrene Interviewer mehr Akquieszenz erhalten als unerfahrene. Gründe dafür können darin liegen, dass erfahrene Interviewer ein höheres Tempo aufweisen. Darüber hinaus stellen erfahrene Interviewer eher ein vertrauliches bzw. harmonisches Verhältnis mit den Befragten her, was wiederum mit erhöhter Akquieszenz einhergeht. Dementsprechend sollte im Interviewertraining insbesondere neutrales Verhalten geschult werden (Olson & Bilgen, 2011).

2.3 Tendenz zur Mitte und zur Milde/Härte

Ein Ausdruck von Satisficing kann die Tendenz zu Antwortkategorien in der Mitte einer Antwortskala sein, unabhängig vom Frageinhalt und unabhängig davon, ob diese Kategorie die tatsächliche Einstellung wiedergibt. Diese Tendenz zur Mitte wird verstärkt, wenn eine Mittelkategorie angeboten wird, was zu einer weniger gründlichen Beantwortung führen kann (Kalton, Robert, & Holt, 1980; Krosnick & Fabrigar, 1997; Schumann & Presser, 1981; Saris & Gallhofer 2007). Für einen Überblick über die Studienlage zu Effekten der Mittelkategorie in Ratingskalen siehe auch SDM Survey Guidelines Artikel „Gestaltung von Ratingskalen in Fragebögen“ (Menold & Bogner, 2015).

Mit der Tendenz zur Milde/ Härte wird die Neigung einiger Befragter beschrieben, die Extremantworten in Ratingskalen – unabhängig vom Inhalt – anzugeben (Greenleaf, 1992), sie ist damit das Gegenteil der Tendenz zur Mitte. Obwohl die Extrempole von Ratingskalen selten genutzt werden, tritt dieses Verhalten bei einigen Befragten unter Umständen auf (Hui & Triandis, 1985).

Gemessen werden kann die Tendenz zu Extremantworten (extreme response style, ERS) als Anteil extremer Antworten in einer Serie von Ratingskalen. Allerdings gibt es hinsichtlich zweier zentraler Fragen keine einheitlichen Befunde: Ist ERS ein stabiles Antwortverhalten bzw. response set? Und: Ist ERS abhängig von persönlichen und demographischen Eigenschaften? (Greenleaf, 1992)

² Ausgeglichene/ balancierte Skalen werden eingesetzt, um Akquieszenz sichtbar zu machen, sie können aber nicht verhindern, dass Akquieszenz die Korrelationen bzw. Faktorstrukturen von Variablen verzerren kann (Rammstedt & Farmer, 2013; Rammstedt, Goldberg, & Borg; 2010).

Die Untersuchung von Greenleaf (1992) kommt zu dem Schluss, dass die Tendenz zu Extremantworten ein über die Zeit bzw. über die Länge eines Surveys stabiles Antwortverhalten ist. Es ist abhängig vom Alter, der Bildung und dem Einkommen der Befragungspersonen, aber nicht vom Geschlecht. Besonders zu Extremantworten neigen ältere Menschen, Menschen mit geringerer Bildung und geringerem Einkommen (Greenleaf, 1992).

Hui und Triandis (1985) kommen zu einem gegenteiligen Ergebnis (Hui & Triandis, 1985): Der Annahme, dass die Tendenz zu Extremantworten stabil über die Zeit ist, stellen sie die Hypothese entgegen, dass Verzerrungen Folge einer Interaktion zwischen Persönlichkeitsfaktoren und situativen Merkmalen wie z.B. zeitlichen Faktoren sind. Diese Annahme können sie bestätigen, Befragte verändern demnach ihr Antwortverhalten im Laufe der Zeit bzw. über die Länge eines Surveys, somit handelt es sich um ein variables Antwortverhalten. Für diese Instabilität bieten sich verschiedene Erklärungen an: Zum einen kann der Versuch, konsistent zu antworten, dazu führen, dass der Befragte am Ende eines Fragebogens Extremantworten gibt, die mit vorher gemachten Angaben korrespondieren. Zum anderen ist die Tendenz zu Extremantworten umso wahrscheinlicher, je mehr ein Befragter mit den Items vertraut ist und je sicherer er sich seiner Einstellung ist (Hui & Triandis, 1985).

Ein weiterer Befund ist die Abhängigkeit der Tendenz zu Extremantworten vom Erhebungsmodus: Telefoninterviews und face-to-face-Interviews sind im Vergleich mit Websurveys oder schriftlichen Befragungen besonders anfällig für die Tendenz zu Extremantworten, aber nicht für extrem negative Antworten. Die festgestellte Tendenz, eher extrem positive Antworten zu geben, könnte mit der Anwesenheit eines (Telefon-) Interviewers zusammenhängen und mit der Abneigung von Befragten, in Gegenwart anderer negative Aussagen zu machen (Ye, Fulton, & Tourangeau, 2011).

3. Einfluss des Interviewers und der Befragungssituation

3.1 Interviewereffekte

Interviewereffekte gehören wie Anwesenheitseffekte zu einem Antwortverhalten, das abhängig von der Situation ist. Die Ursachen liegen in den Verhaltensweisen von Interviewern (Mangione, Fowler, & Louis, 1992) oder in Interviewermerkmalen (Haunberger, 2006). Demographische und sozioökonomische Merkmale des Interviewers können das Antwortverhalten der Befragten beeinflussen, vor allem wenn Items im Fragebogen diese Merkmale thematisieren (Groves, 1989, zit. n. Haunberger, 2006). Die Anwesenheit eines Interviewers hat grundsätzlich positive Auswirkungen: Er kann nachfragen und steht für Nachfragen zur Verfügung, er kann unklare Punkte klären und darauf hinwirken, dass der Befragte die Frage versteht (Mangione et al., 1992). Ein Interviewer kann aber auch Verhaltensweisen zeigen, die negative Auswirkungen haben: Das ist z.B. dann der Fall, wenn Antwortkategorien nicht neutral vorgelesen werden oder nicht die tatsächlich erhaltenen Antworten dokumentiert werden (Mangione et al., 1992).

Zum Zusammenhang zwischen Merkmalen des Fragebogens bzw. den Fragen und der Wahrscheinlichkeit, dass Interviewer das Antwortverhalten der Befragten beeinflussen, gibt es vielfältige Befunde. Relevant ist dabei die Frage, welche Items anfällig für Interviewereffekte sind (Mangione et al., 1992). Als anfällig für Interviewereffekte gelten schwierige Fragen, Meinungsfragen und offene Fragen (Haunberger, 2006). Das Geschlecht und die Bildung des Interviewers können die Beantwortung schwieriger Items beeinflussen: Die Frage nach dem Einkommen wird z.B. eher verweigert, wenn sie von weiblichen Interviewern und von Interviewern mit höherer Bildung gestellt wird (Haunberger, 2006).

Bei den Einstellungsfragen gibt es widersprüchliche Befunde: In einigen Studien kann festgestellt werden, dass Einstellungsfragen nicht anfälliger für Interviewereffekte sind als Faktenfragen (Groves & Magilavy, 1986). In anderen Untersuchungen zeigen sich Zusammenhänge zwischen Interviewergeschlecht und Einstellungsfragen wie z.B. zum Thema Ehe: Weibliche Interviewer erhalten im Gegensatz zu männlichen Interviewern mehr positive und neutrale Antworten als negative (Liu & Stainback, 2013).

Auch offene Fragen ohne vorgegebene Antwortkategorien begünstigen Interviewereffekte (Mangione et al., 1992). Haunberger zeigt einen positiven Effekt zwischen dem Alter und der Bildung der Interviewer und der Beantwortung offener Fragen (Haunberger, 2006). Auch die Anzahl der Nennungen pro offener Frage differiert bei unterschiedlichen Interviewern. Das heißt, die Beantwortung offener Fragen unterscheidet sich zwischen verschiedenen Interviewern nicht unbedingt in der ersten Nennung, aber es kann Unterschiede geben, ob der Befragte eine zweite Antwort nennt; das kann an den unterschiedlichen Verhaltensweisen oder Eigenschaften von Interviewern liegen (Groves & Magilavy, 1986). Bei diesen Auswirkungen handelt es sich eher um geringe Effekte, diese können sich aber innerhalb eines Surveys aufsummieren und statistisch relevant werden, wenn die einzelnen Interviewer viele Interviews durchführen (Groves & Magilavy, 1986; Mangione et al., 1992).

Auch aus der Interaktion von Interviewer- und Befragtenmerkmalen können Interviewereffekte entstehen: Schanz (1981) konnte in seiner Untersuchung einen positiven Zusammenhang zwischen Einstellungs- und Verhaltensmerkmalen von Interviewern und Befragten z.B. in Bezug auf Erziehungsziele oder Alkoholkonsum feststellen (Schanz, 1981). Die Einstellungs- und Verhaltensweisen eines Interviewers können somit, abhängig vom Inhalt einer Frage, Auswirkungen haben: Befragte können von sichtbaren Interviewermerkmalen auf Einstellungen schließen bzw. Interviewer können Einstellungen auch non-verbal ausdrücken (Schanz, 1981). Leistungsmerkmale der Interviewer wie erzielte Rücklaufquote und Produktivität hängen dagegen nicht mit Interviewereffekten zusammen (Groves & Magilavy, 1986).

Verschiedene Studien zeigen, dass das Interviewertraining positive Effekte auf die Datenqualität hat (Haunberger, 2006). Ein Bestandteil des Trainings kann es sein, dass die Interviewer den Fragebogen der Studie selbst ausfüllen. So werden die Einstellungen des Interviewers zum Thema der Erhebung sichtbar gemacht, es können Zusammenhänge zwischen den Einstellungen von Interviewer und Befragten erkannt und mögliche Beeinflussungen identifiziert werden (Haunberger, 2006; Schanz, 1981). Gleichzeitig lernt der Interviewer das Erhebungsinstrument kennen (Haunberger, 2006). Als weitere Implikation ergibt sich, dass es sinnvoll ist, Probing zu trainieren, da der Interviewer gerade durch das Probing das Antwortverhalten beeinflussen kann. Probing tritt besonders häufig bei schwierigen und offenen Fragen auf. Ein wichtiger Ansatzpunkt besteht also darin, die Fragen so zu formulieren, dass sie möglichst wenig Probing erfordern. Pretests können bei der Verbesserung der Itemformulierung helfen (Mangione et al., 1992). Detaillierte Informationen zum Interviewertraining enthält das Kapitel Interviewer Qualifikation und Training der SDM Survey Guidelines (Stiegler & Biedinger, 2014).

3.2 Anwesenheit dritter Personen

Anwesenheitseffekte können auftreten, wenn in einem Interview dritte Personen anwesend sind und das Antwortverhalten der Befragungsperson von dieser Anwesenheit Dritter abhängig ist (Lander, 2000). Anwesende können Partner, Kinder, andere Verwandte, aber auch Freunde oder Nachbarn sein, in den allermeisten Fällen sind es die Ehepartner (Hartmann, 1994).

Die meisten Studien, die die Effekte der Anwesenheit Dritter untersuchen, berichten positive Auswirkungen auf die Beantwortung von Faktenfragen: Ehepartner können kontrollieren, ob zutreffende Angaben gemacht werden (Lander, 2000) und somit dazu beitragen, die Datenqualität zu

verbessern (Zipp & Toth, 2002; Aquilino, 1993). Die Anwesenheit des Partners kann den Befragten helfen, sich zu erinnern (Hartmann, 1994).

Daneben werden jedoch auch Auswirkungen in Bezug auf die Validität der erhobenen Daten durch veränderte Verhaltenserwartungen aller Beteiligten, also Interviewer, Befragter und anwesender Dritter, diskutiert (Lander, 2000). Die Anwesenheit Dritter hat (als methodisches Problem) Einfluss auf die Datenqualität, weil sie soziale Konsequenzen für den Befragten haben kann (Hartmann, 1994) – insofern sind Anwesenheitseffekte als Einfluss sozialer Erwünschtheit sowohl gegenüber dem Interviewer als auch gegenüber der dritten Person interpretierbar (Lander, 2000). Wenn bei Befragungen der Ehepartner anwesend ist, sind Unterschiede in den Daten feststellbar. Die Anwesenheit des Partners kann die Antworten auf partnerschaftsbezogene Fragen beeinflussen (Lander, 2000). Insgesamt sind jedoch ein geringer Einfluss der Anwesenheit des Partners nachweisbar und wenige Belege für soziale Erwünschtheit vorhanden (Zipp & Toth, 2002). Die Anwesenheit des Ehepartners führt zu einer erhöhten Einigkeit zwischen den Partnern bei Einstellungs- und Verhaltensfragen, wenn beide befragt werden (Zipp & Toth, 2002) und zu einer Zunahme traditioneller Antworten; diese Zunahme kann nicht nur auf der Anwesenheit des Ehepartners zurückgeführt werden, sondern die Anwesenheit des Ehepartners kann ein Ausdruck einer traditionelleren Partnerschaft sein (Lander, 2000). Neuere Analysen zeigen also keine systematischen Einflüsse der Anwesenheit Dritter, auch wenn bei einzelnen Aspekten uneinheitliche Befunde berichtet werden (Lander, 2000).

Viele Studien beweisen, dass die Anwesenheit eines Partners nicht zufällig ist, sondern von Merkmalen des anwesenden Ehepartners und des Befragten abhängt. Dabei handelt es sich um Merkmale, die nicht nur die Anwesenheit begründen, sondern auch Einfluss auf das Antwortverhalten haben können (Hartmann, 1994). Bei den resultierenden unterschiedlichen Antwortverhalten handelt es sich dann nicht zwangsläufig um Anwesenheitseffekte und Verzerrungen (Hartmann, 1994), eben weil die Anwesenheit des Partners und die Beantwortung der Fragen von den gleichen Merkmalen beeinflusst werden (Aquilino, 1993). Die Anwesenheit eines Partners kann neben strukturellen Gründen (z.B. Nichterwerbstätigkeit) auch motivationale Hintergründe eines oder beider Partner haben, die auch bei den partnerschaftsbezogenen Fragen relevant werden können (Lander, 2000).

Für die Forschungspraxis ergeben sich einige Implikationen: Es hat sich gezeigt, dass mit zunehmender Interviewdauer eine Anwesenheit des Ehepartners wahrscheinlicher wird. Wenn man also den Befragten alleine interviewen möchte, sollten die Interviews eher kurz sein (Hartmann, 1994). Die Interviewer sollten angehalten werden, Informationen über die Anwesenheit Dritter zu sammeln (Aquilino, 1993): Es ist zwar oft bekannt, ob Dritte anwesend waren, aber nicht deren Gründe und die Motive für die Anwesenheit (Hartmann, 1994). Insgesamt ist festzustellen, dass die Anwesenheit des Partners kaum zu Antwortverzerrungen führt und den Befragten helfen kann, sich zu erinnern. Dennoch sollte der Forscher sich des Kontexts des Interviews bewusst sein, wenn die Interviewer-Befragten-Interaktion um anwesende Dritte erweitert wird (Zipp & Toth, 2002).

4. Einfluss des Fragebogendesigns

4.1 Effekte der Fragereihenfolge

Nicht nur die Formulierung von Fragen, sondern auch die Reihenfolge in der sie gestellt werden, kann Einfluss auf ihre Interpretation und Bewertung von Seiten der Befragten ausüben, da vorangehende Fragen den Kontext für nachfolgende Fragen bilden (vgl. u.a. Tourangeau et al., 2000; Moore, 2002; Krosnick & Presser, 2010). Dabei können sich Befragte bei der Beantwortung von Folgefragen an vorhergehenden Fragen bzw. zuvor gegebenen Antworten orientieren (s.g. Ausstrahlungseffekte oder

Halo-Effekte), wodurch eine systematische Verzerrung der Antworten zustande kommt (Schnell, Hill, & Esser, 2011; Schnell, 2014).

Vorangehende Fragen können das Abrufen bestimmter Informationen auslösen, welche dann für nachfolgende Fragen zu ähnlichen Themen sehr leicht für den Befragten zugänglich sind. Dies führt dazu, dass Befragte ihre Antworten zu späteren Fragen wahrscheinlich auf dieselben Überlegungen aufbauen wie vorangehende Fragen zu ähnlichen Themen. Wird ein Befragter beispielsweise danach gefragt, wie zufrieden er mit seiner Ehe ist, erhöht dies die Chance, dass der Befragte die Ehe auch bei der Beantwortung der Frage nach der Zufriedenheit mit dem Leben allgemein einbezieht. Dieser Effekt wird Assimilationseffekt genannt. Im Gegensatz zu Assimilationseffekten stehen Kontrasteffekte: Antworten zu späteren Fragen werden hierbei in die entgegengesetzte Richtung gelenkt wie die zu vorangehenden Fragen. Bei der Beantwortung von späteren Fragen können Befragte absichtlich die Information aus dem Antwortprozess ausschließen, die sie bei vorangehenden Fragen genutzt haben, um Wiederholungen zu vermeiden (Schwarz & Bless, 1992; Mason, Carlson, & Tourangeau, 1994).

Um das Auftreten von Fragereihenfolgeeffekte zu identifizieren, kann die Reihenfolge der Fragen im Pretest oder in einem Split-Ballot-Design variiert werden. Dadurch erhöht sich jedoch auch der Aufwand in der Fragebogenkonstruktion.

4.2 Reihenfolge der Antwortvorgaben

Die Anordnung der Antwortvorgaben (Response Order) in einer Liste oder Ratingskala kann Einfluss darauf nehmen, welche Antwort ein Befragter auswählt. Bei einer visuellen Präsentation der Antwortalternativen, wie auf Showcards oder in selbstadministrierten Befragungen wie Web Surveys, tendieren Befragte dazu eine der oberen Alternativen zu wählen (Primacy-Effekt). Anders verhält es sich bei mündlich vorgebrachten Antwortalternativen, zum Beispiel in (Telefon-) Interviews, bei denen Befragte eher aus den zuletzt genannten Alternativen wählen (Recency-Effekt) (Krosnick & Alwin, 1987; Krosnick, 1991; Schwarz, Hippler, & Noelle-Neumann, 1992; Bishop & Smith, 2001).

Zur Erklärung für die Entstehung von Response-Order-Effekten haben sich zwei theoretische Modelle bewährt: das Response-Order Elaboration Model (Sudman, Bradburn, & Schwarz, 1996) und die Satisficing-Theorie (Krosnick et al., 1996). Nach dem Response-Order Elaboration Model ist die Entstehung von Response-Order Effekten im Wesentlichen abhängig davon, welche Chance ein Befragter hat, über die Inhalte der einzelnen Antwortalternativen nachzudenken. Diese Chance wird von drei Faktoren bestimmt: (1) die serielle Position in welcher eine Alternative präsentiert wird (am Anfang, in der Mitte oder am Ende); (2) der Modus der Präsentation (mündlich durch Vorlesen oder visuell in selbst-administrierten Befragungen oder mittels Showcards in Interviews) und (3) die Plausibilität einer Antwortalternative. Gemäß den Modellannahmen werden dann für visuell präsentierte Antwortformate Primacy-Effekte erwartet, weil Befragte eine bessere Chance haben über die ersten Alternativen nachzudenken. Dies trifft nicht zu, wenn die ersten Alternativen dem Befragten nicht plausibel erscheinen. Für diesen Fall werden auch bei visuellen Formaten Recency-Effekte vorhergesagt. Für mündlich vorgebrachte Antwortformate wird das Auftreten von Recency-Effekten erwartet, weil Befragte beim Vorlesen der Alternativen kaum eine Chance haben, die erstgenannten Alternativen kognitiv zu verarbeiten, da bereits die Nennung der nächsten Alternative diesen Prozess wieder unterbricht. Auch in diesem Fall wird angenommen, dass bei nicht plausibel erscheinenden Alternativen auch bei mündlicher Präsentation Primacy-Effekte auftreten (Bishop & Smith 2001).

Aus Perspektive der Satisficing-Theorie entstehen Response-Order-Effekte, wenn Befragte aus Gründen geringer Motivation, geringer kognitiver Fähigkeiten oder einer hohen Aufgabenschwierigkeit den kognitiven Prozess der Fragebeantwortung vereinfachen wollen. Dann nutzen sie die erste akzeptabel erscheinende Antwortalternative zur Antwortgabe, ohne den weiteren Alternativen besondere Aufmerksamkeit zu schenken (Krosnick & Alwin, 1987; Krosnick, 1991). Die erwarteten Effektrichtungen

sind dann wie im Response-Order Elaboration Model bei visuell präsentierten Formaten Primacy-Effekte und bei mündlich präsentierten Formaten Recency-Effekte (Bishop & Smith, 2001).

Galesic, Tourangeau, Couper & Conrad (2008) analysieren mittels der Eye-Tracking-Methode die Blickbewegungen und Fixationszeiten von Befragten bei der Beantwortung eines Web Fragebogens. Dabei konnten zwei Antwortverhalten bei Befragten identifiziert werden, die als Ausdruck von schwachem und starkem Satisficing beurteilt werden können. Einige Befragte wählten Alternativen aus der ersten Hälfte der Liste, ohne eine der unteren Alternativen überhaupt zu fixieren (starkes Satisficing). Andere Befragte zeigten unabhängig von den Inhalten der Antwortalternativen, signifikant längere Fixationszeiten für die obere Listenhälfte im Vergleich zur unteren, was auf einen gründlicheren kognitiven Prozess bei der Verarbeitung der oberen Alternativen hinweist (schwaches Satisficing). Des Weiteren verdeutlicht der Stand der Forschung, dass Primacy-Effekte umso wahrscheinlicher auftreten, umso länger die Liste an Antwortalternativen ist (Galesic et al., 2008), umso geringer die kognitiven Fähigkeit von Befragten sind (Krosnick & Alwin, 1987) und umso höher das Alter der Befragten ist, Bildung nimmt keinen Einfluss auf Response-Order Effekte (Knäuper, 1999).

Anhand von Daten aus 548 Experimenten - durchgeführt in Telefonstudien - untersuchten Holbrook, Krosnick, Moore & Tourangeau (2007) welche Merkmale von Fragen und Befragten das Entstehen von Recency-Effekten begünstigen. Es zeigt sich, dass Recency-Effekte eher bei komplizierten Fragen auftreten. Dieser Effekt wird für weniger gebildete Befragte verstärkt, wenn die Fragen komplizierter sind, sich am Ende des Fragebogens befinden und wenn die Antwortalternativen als Sätze und nicht als Schlagworte/ Phrasen formuliert sind. Wurden Befragten implizit oder explizit in der Fragestellung aufgefordert, ihre Antwort erst zu bilden, nachdem alle Alternativen vorgelesen wurden, kam es trotzdem vermehrt zu Recency-Effekten (Holbrook et al., 2007).

Um Response-Order-Effekten entgegenzuwirken wird empfohlen, bei der Entwicklung eines Frageinstrumentes - wann immer möglich - lange Antwortlisten zu vermeiden. Ansonsten können noch folgende Maßnahmen in Betracht gezogen werden: Bei Listen mit ungeordneten Antwortalternativen können diese einer Rotation unterzogen werden, damit immer andere Alternativen an oberer / unterer Stelle erscheinen und ein Bias in den Daten vermieden wird (Holbrook et al., 2007). Dieses Verfahren ist bei computerunterstützten Befragungen einfach umsetzbar, bei anderen Befragungsformen und vor allem bei Papierbefragungen, sehr aufwendig und kostenintensiv. Auch ist das Problem nicht behoben, dass weiterhin der einzelne Befragte anstatt seiner tatsächlichen Einstellung Antworten gibt, die von Response-Order-Effekten beeinflusst wurden. Daher ist eine weitere Maßnahme, die Motivation der Befragten zu erhöhen, um einem Satisficing-Verhalten vorzubeugen und dem Entstehen von Response-Order-Effekten entgegenzuwirken. Dies kann etwa durch Instruktionen erfolgen, wie zum Beispiel, dass nun eine schwierige Frage folgt, für deren Beantwortung viel Konzentration erforderlich ist (Krosnick & Alwin, 1987). Die generelle Empfehlung, Fragen so einfach wie möglich zu formulieren und auf ungewöhnliche und abstrakte Begriffe zu verzichten, ist auch im Hinblick auf Response-Order-Effekte auszusprechen. Denn mit einfach gehaltenen Fragen wird auch die kognitive Belastung der Befragten so gering wie möglich gehalten was wiederum dem Entstehen von Satisficing entgegenwirkt (Holbrook et al., 2007). Des Weiteren sind Forced-Choice den Check-All Antwortformaten vorzuziehen, da für diese weniger Primacy-Effekte und eine tiefere Informationsverarbeitung der Befragten beobachtet wurde (Smyth, Dillman, Christian, & Stern, 2006). Schließlich ist noch zu überlegen, ob alternativ zu einer Liste mit Antwortalternativen nicht Ratingskalen zur Bewertung jeder einzelnen Alternative eingesetzt werden können. Hier ist aber auch abzuwägen, ob dadurch nicht wiederum der Fragebogen so stark an Umfang zunimmt, dass andere negative Effekte auftreten. Weitere Empfehlungen zur Gestaltung von Antwortformaten sind dem SDM Survey Guidelines Artikel „Frageformulierung“ (Lenzner & Menold, 2015) zu entnehmen.

Literaturverzeichnis

- Aquilino, W. S. (1993). Effects of spouse presence during the interview on survey responses concerning marriage. *Public Opinion Quarterly*, 57, 358-376.
- Belli, R. F., Traugott, M. W., & Beckmann, M. N. (2001). What leads to voting overreports? Contrasts of overreporters to validated votes and admitted nonvoters in the American Election Studies. *Journal of Official Statistics*, 17, 479-498.
- Billiet, J. B., & Davidov, E. (2008). Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. *Sociological Methods & Research*, 36, 542-562.
- Bishop, G., & Smith, A. (2001). Response-order effects and the early Gallup split-ballots. *Public Opinion Quarterly*, 65, 479-505.
- Burden, B. C. (2008). The social roots of the partisan gender gap. *Public Opinion Quarterly*, 72(1), 55-75.
- Cannell, C. F., Miller, P. V., & Oksenberg, L. (1981). Research on interviewing techniques. In: S. Leinhardt (Ed.), *Social Methodology*. San Francisco: Jossey-Bass Publishers.
- Galesic, M., Tourangeau, R., Couper, M., & Conrad, F. (2008). Eye tracking data. New insights on response order effects and other cognitive shortcuts in survey responding. *Public Opinion Quarterly*, 72(5), 892-913.
- Greenleaf, E. A. (1992). Measuring extreme response style. *Public Opinion Quarterly*, 56, 328-351.
- Groves, R. M., & Magilavy, L. J. (1986). Measuring and explaining interviewer effects in centralized telephone surveys. *Public Opinion Quarterly*, 50(2), 251-266.
- Groves, R. M., Fowler, J. F., Couper, M., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology*. Hoboken: John Wiley & Sons.
- Hartmann, P. (1994). Interviewing when the spouse is present. *International Journal of Public Opinion Research*, 6(3), 298-306.
- Haunberger, S. (2006). Das standardisierte Interview als soziale Interaktion: Interviewereffekte in der Umfrageforschung. *ZA-Information*, 58, 23-46.
- Holbrook, A. L., & Krosnick, J. A. (2010). Social desirability bias in voter turnout reports. Tests using the item count technique. *Public Opinion Quarterly*, 74(1), 37-67.
- Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephon versus face-to-face interviewing of national probability samples with long questionnaires. *Public Opinion Quarterly*, 67, 79-125.
- Holbrook, A. L., Krosnick, J. A., Moore, D., & Tourangeau, R. (2007). Response order effects in dichotomous categorical questions presented orally. The impact of question and respondent attributes. *Public Opinion Quarterly*, 71(3), 325-348.
- Hui, C. H., & Triandis, H. C. (1985). The instability of response sets. *Public Opinion Quarterly*, 49, 253-260.
- Kahn, R. L., & Cannell, C. F. (1957). *The Dynamics of Interviewing: Theory, Technique, and Cases*. New York: Wiley.
- Kemper, C. J., Beierlein, C., Bensch, D., Kovaleva, A., & Rammstedt, B. (2012). Eine Kurzsкала zur Erfassung des Gamma-Faktors sozial erwünschten Antwortverhaltens. Die Kurzsкала Soziale Erwünschtheit-Gamma (KSE-G). *GESIS-Working Papers*, Nr. 2012|25. Köln: GESIS- Leibniz Institut für Sozialwissenschaften.

- Knäuper, B. (1999). The impact of age and education on response order effects in attitude measurement. *Public Opinion Quarterly*, 63, 347-370.
- Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and web surveys. *Public Opinion Quarterly*, 72(5), 847-865.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213-236.
- Krosnick, J. A. (1999a). Maximizing questionnaire quality. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of Political Attitudes* (pp. 37-57). San Diego: Academic Press.
- Krosnick, J. A. (1999b). Survey research. *Annual Review of Psychology*, 50, 537-567.
- Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*, 51(2), 201-219.
- Krosnick, J. A., & Fabrigar, L. R. (1997). Designing rating scales for effective measurement in surveys. In L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey Measurement and Process Quality* (pp. 141-164). New York: Wiley-Interscience.
- Krosnick, J. A., & Presser, S. (2010). Question and questionnaire design. In P. V. Marsden, & J. D. Wright (Eds.), *Handbook of Survey Research* (pp. 263-313). Bingley: Emerald Group Publishing.
- Krosnick, J. A., Narayan, S. S., & Smith, W. R. (1996). Satisficing in surveys: Initial evidence. In M. T. Braverman, & J. K. Slater (Eds.), *Advances in Survey Research* (pp. 29-44). San Francisco: Jossey-Bass.
- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality and Quantity*, 47, 2025-2047.
- Lander, B. (2000): Anwesenheitseffekte im Wandel. Eine Sekundäranalyse zur Anwesenheit des Partners im Interview anhand des ALLBUS 1980 bis 1998. *Zeitschrift für Soziologie*, 29(3), 227-239.
- Lenzner, T. & Menold, N. (2015). Frageformulierung. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (SDM Survey Guidelines). DOI: 10.12465/sdm-sg_017
- Liu, M., & Stainback, K. (2013). Interviewer gender effects on survey responses to marriage related questions. *Public Opinion Quarterly*, 77(2), 606-618.
- Mangione, T. W., Fowler, F. J., & Louis, T. A. (1992). Question characteristics and interviewer effects. *Journal of Official Statistics*, 8(3), 293-307.
- Mason, R., Carlson, J. E., & Tourangeau, R. (1994). Contrast effects and subtraction in part-whole questions. *Public Opinion Quarterly*, 58, 569-578.
- McBride, L. & Moran, G. (1967). Double agreement as a function of item ambiguity and susceptibility to demand implications of the psychological situation. *Journal of Personality and Social Psychology*, 6, 115-118.
- Menold, N. & Bogner, K. (2015). Gestaltung von Ratingskalen in Fragebögen. Mannheim, GESIS – Leibniz Institut für Sozialwissenschaften (SDM Survey Guidelines). DOI:10.15465/sdm-sg_015
- Moore, D. W. (2002). Measuring new types of question-order effects. *Public Opinion Quarterly*, 66(1), 80-91.
- Olson, K., & Bilgen, I. (2011). The role of interviewer experience on acquiescence. *Public Opinion Quarterly*, 75(1), 99-114.
- Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49-69). Mahwah NY: Erlbaum.

- Preisendörfer, P., & Wolter, F. (2014). Who is telling the truth? A validation study on determinants of response behavior in surveys. *Public Opinion Quarterly*, 78(1), 126-146.
- Rammstedt, B. & Farmer, R. F. (2013). The impact of acquiescence on the evaluation of personality structure. *Psychological Assessment*, 25, 1137-1145.
- Rammstedt, B., Goldberg, L. R. & Borg, I. (2010). The measurement equivalence of Big-Five factor markers for persons with different levels of education. *Journal of Research in Personality*, 44, 53-61.
- Rasinski, K. A. (1989). The effect of question wording on public support for government spending. *Public Opinion Quarterly*, 53, 388-394.
- Schanz, V. (1981). Interviewereffekte: zusammenfassende Darstellung von Arbeiten, die im Rahmen zweier von ZUMA betreuter Projekte entstanden sind. *ZUMA Nachrichten*, 5(9), 36-46.
- Schnell, R. (2014). *Survey-Interviews. Methoden standardisierter Befragungen*. Wiesbaden, VS Verlag für Sozialwissenschaften.
- Schnell, R., Hill, P. B., & Esser, E. (2011). *Methoden der empirischen Sozialforschung*. München, Oldenbourg Verlag.
- Schwarz, N. & Bless, H. (1992). Assimilation and contrast effects in attitude measurement: An inclusion/exclusion model. *Advances of Consumer Research*, 19, 72-77.
- Schwarz, N., Hippler, H.-J., & Noelle-Neumann, E. (1992). A cognitive model of response order effects in survey measurement. In N. Schwarz, & S. Sudman (Eds.), *Context Effects in Social and Psychological Research*. New York: 187-201.
- Smyth, J. D., Dillman, D. A., Christian, L. M., & Stern, M. J. (2006). Comparing check-all and forced-choice question formats in web surveys *Public Opinion Quarterly*, 70, 66-77.
- Stiegler, A. & Biedinger, N. (2015). Interviewer Qualifikation und Training. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (SDM Survey Guidelines). DOI: 10.15465/sdm-sg_013
- Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass.
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5), 859-883.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Trott, D. M. & Jackson, D. N. (1967). An experimental analysis of acquiescence. *Journal of Experimental Research in Personality*, 2, 278-288.
- Ye, C., Fulton, J., & Tourangeau, R. (2011). Research synthesis. More positive or more extreme? A meta-analysis of mode differences in response choice. *Public Opinion Quarterly*, 75(2), 349-365.
- Zipp, J. F., & Toth, J. (2002). She said, He said, They said. The impact of spousal presence in survey research. *Public Opinion Quarterly*, 66, 177-208.