

## *Exkurs: geschlossenes versus offenes Frageformat*

Offene Fragen unterscheiden sich auf den ersten Blick von geschlossenen durch die Antwortmöglichkeiten. Während offene Fragen in dieser Hinsicht keinen Beschränkungen unterworfen sind, ist die Anzahl von Möglichkeiten zur Antwort bei geschlossenen Fragen bereits vor der Befragung festgelegt und damit dem Befragten vorgegeben. Zu den geschlossenen Fragen zählen auch Ratingskalen, bei denen das Zutreffen von oder die Zustimmung zu Aussagen anhand einer vorgegebenen numerischen Skala deklariert wird.

Nach Tränkle (1983) stellen die Frageformen eine jeweils andere kognitive Anforderung an den Befragten. Bei offenen Fragen soll er frei reproduzieren, während er bei geschlossenen Fragen eine Wiedererkennungseistung erbringen muß, um die zutreffende Antwortmöglichkeit auswählen zu können. In diesem Sinne besteht auch eine auf Eysenck zurückgehende Differenzierung in einen kreativen und einen selektiven Fragetyp.

Obwohl verschiedene kognitive Vorgänge bei der Beantwortung angenommen werden, bevorzugt man in den Verhaltens- und Sozialwissenschaften den geschlossenen Fragentyp deutlich gegenüber dem offenen Frageformat. Wie läßt sich diese Präferenz begründen? Etwa durch Validitätsvorteile auf Seiten der geschlossenen Frageform? Diese Frage ist trotz einer in der Literatur vorzufindenden Debatte nicht einfach zu beantworten. Es gibt wenig systematische Forschung zu dieser Fragestellung, sondern nur vereinzelte Befunde, unter denen jedoch vor allem die Arbeiten von Belson und Duncan (1962) sowie Schuman und Presser (1996) hervorzuheben sind.

Converse (1984) sieht für die Dominanz geschlossener Fragen im heutigen Wissenschaftsalltag vor allem historische und institutionelle Gründe. Während des Zweiten Weltkrieges erhielten in den USA zwei konkurrierende Gruppen von Forschungsinstitutionen staatliche Aufträge, der Regierung mit wissenschaftlichen Methoden Übersicht über die Stimmung in der Bevölkerung zu verschaffen. Auf der einen Seite standen dabei privatwirtschaftlich organisierte und soziologisch orientierte Forschungsgruppen, deren Gegenstand Umfragen gewesen sind (polls - einer der heute noch bekannten Leiter einer dieser Gruppen war Gallup). Die andere Seite nahmen Forscher mit universitärem Hintergrund aus der sozialpsychologischen Meinungsforschung (surveys) ein. Während die kommerziellen soziologischen Institutionen in den Umfragen vor allem geschlossene Fragen verwendeten, benutzten ihre akademischen Konkurrenten vor allem offene Fragen. Erstaunlicherweise ist Likert einer der Wortführer letzterer Forschungsrichtung gewesen, obwohl sein Name heute für die von ihm eingeführte Ratingskala mit Normalverteilungseigenschaft steht.

Die privaten Anbieter setzten sich gegen ihre akademische Konkurrenz durch, da sie in der Lage waren, unter Anwerbung von ungeschulten Hilfskräften in kürzerer Zeit größere Stichprobenumfänge kostengünstiger zu realisieren. Die Surveyforscher mußten ihre bereits hochqualifizierten Interviewer noch zusätzlich für jedes Projekt schulen, um dann bei entsprechen höheren Kosten nur einen geringeren Datendurchsatz bei höherem Zeitaufwand zu erzielen. Die daraus resultierenden Nachteile bei der Bewerbung um Forschungsaufträge versuchten sie in einer öffentlichen methodologischen Kontroverse auszugleichen. Wegen des großen Gewichtes der Forschungsergebnisse bei schwerwiegenden gesellschaftliche Veränderungen gab es dazu hinreichend Anlaß.

So fanden die Umfrageforscher große Zustimmung dafür, in den USA befindliche Japaner während des Krieges in Internierungslagern unter Bewachung zu stellen. Die Meinungsforscher führten diese Ergebnis auf die geschlossene Formulierung der entsprechenden Antwortalternativen zurück, denn von ihnen parallel im offenen Format Befragten empfanden die Japaner, die sich im Land befanden, keineswegs als Bedrohung und nur eine Minderheit von ihnen sprach sich für eine Internierung aus.

Doch trotz methodischer Einwände gegen geschlossenes Fragen sprachen die ökonomischen Eckdaten gegen eine Meinungsforschung im offenen Antwortformat. Die entsprechenden Forschungsstrukturen wurden noch während des Krieges abgewickelt. Converse kommt demzufolge zu dem Schluß: „The open/closed debate was shaped in good part by institutional needs and capacities, and by ideology *about* research, remaining largely untouched *by* research.“ und seiner Überzeugung nach beruhen Entscheidungen über das zu wählende Frageformat deshalb mangels empirisch gestützter Erkenntnisse auf „enlightend common sense, in-house experience, and practical constraints“ (1984, S. 279).

Eine Verbesserung im Erkenntnisstand hinsichtlich eines Effektes aufgrund des Formates der Frage brachte die Untersuchungen von Belson und Duncan (1962). In einer Gruppe von Studien werden die Antworten von offene und geschlossenen Fragen nach der Zeitungs- und Zeitschriftenlektüre des vorherigen Tages verglichen. Das Thema in einer zweiten Staffel ist die Abhängigkeit der Angaben zu den gesehenen Fernsehsendungen des Vortages vom Frageformat Die Autoren weisen auf stärkere Unterschiede zwischen den Antworten hin, als bei den ja inhaltlich gleichen Fragen zu erwarten war. Insbesondere lag die Besetzungsrate derjenigen Kategorien, die im geschlossenen Format vorgegeben waren, bei den offen gestellten Fragen fast durchgängig um 20% niedriger. Als Kehrseite dieses Befundes liegt die Anzahl der Antwortkategorien im offenen Format meist deutlich über der Anzahl von vorgegebenen Antwortmöglichkeiten der geschlossenen Frage.

Bei offenen Fragen werden vor allem diejenigen Sachverhalte angegeben, mit denen die Befragten vertraut sind. Zufällige oder neue Begebenheiten werden in diesem Format deutlich weniger häufig berichtet. Sind Sachverhalte jedoch so selbstverständlich und alltäglich wie beispielsweise das Sehen der Abendnachrichten, werden diese auch etwas seltener als im geschlossenen Format berichtet. Eine zusätzliche offene Antwortkategorie im geschlossenen Antwortformat wie etwa „Anderes ...“ wird von den Befragten praktisch nicht benutzt („depressing effect of check-list“ Belson & Duncan, 1962, S. 127). Offensichtlich wird die Liste von Antwortalternativen durch die Befragten für erschöpfend und die Abarbeitung dieser für hinreichend für die Beantwortung einer Frage gehalten.

Belson und Duncan betonen, daß sie nur Unterschiede zwischen beiden Formaten aufzeigen, aber kein abschließendes Urteil über die Gültigkeit einer der beiden Befragungsformen treffen können. Trotzdem halten sie die Unterschiede für so gravierend, daß in einem gegebenem Kontext eigentlich nur eines der beiden Formate valide Daten liefern kann. Im Anwendungsfall sollte deshalb stets geprüft werden, ob eine ungünstige Wahl des Antwortformates die Validität gefährdet.

Während Belson und Duncan (1962) das Verhalten des Vortages zum Gegenstand ihrer Studie machten, untersuchten Schuman und Presser (1997) den Effekt des Antwortformates bei der Einstellungsmessung. Dabei haben die Autoren ihre Fragestellungen in mehrere Kampagnen von Surveys mit repräsentativen Stichproben einbinden können (Detroit Area Study, Survey Research Center National Omnibus Survey und National Opinion Research Center General Social Survey). Der Effekt des Antwortformates wurde durch eine Split-Ballot-Technik in drei Einstellungsbereichen systematisch untersucht: das wichtigste anstehende nationale Problem, Arbeitswerte und Werte der Kindererziehung.

Den Ausgangspunkt bilden zwei Annahmen bezüglich der Vorteile des offenen Frageform:

- 1) vorgefertigten Antworten der geschlossenen Fragen kann es möglicherweise an Verständlichkeit oder inhaltlicher Relevanz mangeln und
- 2) durch die vorgegebenen Antworten können die Befragten in ihrer Antwort beeinflusst sein.

Es sei angemerkt, daß Schuman und Presser genau genommen keine eigentlichen Vorteile der offenen Frageform ansprechen sondern vielmehr Nachteile bei geschlossenen Fragen. Offensichtlich stellt nach Meinung der Autoren ein Nachteil auf der einen Seite einen Vorteil bei der anderen dar.

Als Maß für die Auswirkung auf das Antwortverhalten dienen Unterschiede in den Randverteilungen der jeweiligen Antworten unter den beiden Bedingungen des Antwortformates sowie

Interaktionen der Antworten mit den Merkmalen Geschlecht und Bildungsgrad. Bei einer entscheidenden Abweichung der Randverteilungen zwischen den Gruppen kann ein wie oben beschriebener Nachteil der geschlossenen Fragen nicht ausgeschlossen werden.

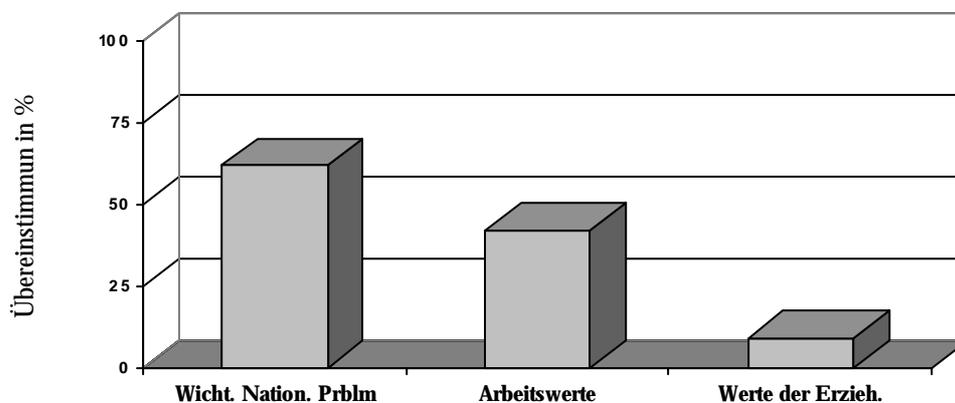


Abbildung 6: Anteil der vorgegebenen Antwortkategorien an den Antworten im offener Frageform in den drei Einstellungsbereichen bei Schuman und Presser (1997): wichtigstes anstehendes nationales Problem, Arbeitswerte und Werte der Kindererziehung.

Bereits bei der Betrachtung der Antwortkategorien wird augenfällig, daß die beiden Frageformate nur schwer vergleichbare Antwortmengen hervorbringen. Geht man von den Antwortalternativen des geschlossenen Formates aus, so liegt der Anteil gleichlautender Antworten im offenen Format bei keinem der Einstellungsbereiche über 62%. Mindestens eine von drei Antworten des offenen Formates ist also nicht durch die vorgegeben Menge von Antwortalternativen der geschlossenen Fragen abgedeckt. Dramatisch stellt sich der Unterschied bezüglich der geäußerten Werte der Kindererziehung dar. Hier liegt die Übereinstimmung von Antworten der beiden Frageformen unter 10% (Abbildung 6). In diesem extremen Fall ist mehr von einem qualitativen als von einem quantitativen Unterschied zwischen den Frageformen zu sprechen. Die vorgegebenen Antwortmöglichkeiten geben nur einem Bruchteil die frei geäußerten Wertvorstellungen bei der Kindererziehung wieder. Es spricht alles dafür, daß sich hier beide Nachteile der geschlossenen Frage besonders ungünstig auswirken.

Allgemein fallen die Übereinstimmungen jedoch zwischen den einzelnen Antwortkategorien sehr unterschiedlich aus. Wenn der relative Anteil von Antworten als Indikator für die Bedeutung des Inhalts einer Antwortkategorie zu verstehen ist, erhalten beispielsweise bei der Frage nach den wichtigsten anstehenden nationalen Problemen die Kategorien Arbeitslosigkeit und Inflation unabhängig vom Frageformat die gleiche Bedeutung. Eine deutlich geringere Bedeutung wird im offenen Format Kriminalität und Gewalt sowie der Abnahme des Vertrauens in die Regierung zugeschrieben. Diese Verschiebungen sind Ausdruck der erheblichen Unterschiede in den Randverteilungen der Nennungshäufigkeiten von Antwortkategorien zwischen den Befragungsgruppen.

Darüber hinaus bestehen Unterschiede im Antwortverhalten zwischen den Geschlechtern und den Bildungsstufen. Dabei sind die Unterschiede zwischen Männern und Frauen bei offenen und geschlossenen Fragen unter Umständen wie beim Arbeitswert Bezahlung gegenläufig (Abbildung 7).

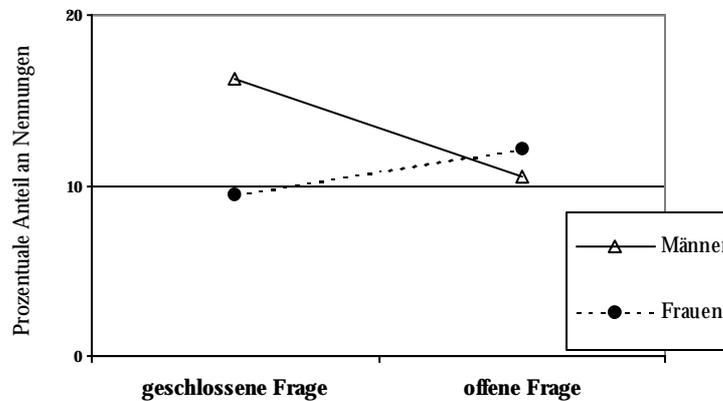


Abbildung 7: Abhängigkeit der relativen Häufigkeit der Antworten zum Arbeitswert Bezahlung von der Frageform und dem Geschlecht (nach Schuman & Presser, 1997, S.92).

Zwischen den Befragten mit Hochschulbildung ergeben sich weniger starke Unterschiede in der Übereinstimmung von offenen und geschlossenen Antworten als bei Befragten ohne Hochschulabschluß..

Die vorgefundenen qualitativen Unterschiede zwischen den Frageformen führen Schuman und Presser zu der Einsicht: „Not every existing closed attitude question can be asked in a parallel open form. In fact, ... , we found most closed items unsuitable for transformation into open form.“(1997, S. 107). Aus diesem Grund wird der Annahme einer “form-resistant correlation“ von Inhalten, die in verschiedenen Frageformaten erhoben werden kann, verworfen.

Schuman und Presser empfehlen deshalb: „Important questions should be developed first by using an open question form in a sufficiently large pilot study, then in most cases turned into closed questions on the basis of the main open codes needed to summarize the open data“ (1997, S. 312).

Die bisher dargestellten Arbeiten von Belson und Duncan (1962) sowie von Schuman und Presser (1997) wiesen erhebliche Effekte des Antwortformates nach. Ob diese jedoch die Validität der Antworten berühren, konnte nicht gezeigt werden. Das liegt in erster Linie daran, daß die Ausprägung der zu messenden Merkmale unbekannt gewesen ist und somit eine Außenkriterium zur Überprüfung der Validität fehlte. Birenbaum und Tazsuoka (1987) benutzten dahingegen das wohldefinierte Material eines mathematischen Leistungstests, um den Effekt des Fragenformates bei Schülern im Alter von ca. 14 Jahren zu untersuchen. Die dabei verwendeten Aufgaben zur Addition von Brüchen sind nach einer ausführlichen Fehleranalyse im

Anschluß an eine Pilotstudie so ausgewählt, daß mit ihrer Hilfe übliche Fehler beim Lösen dieses Aufgabentyps erfaßt werden können. Im Mittelpunkt des Tests steht somit nicht nur die Leistung bei der Addition von Brüchen, sondern vielmehr auch die Identifikation der eventuell dabei begangenen Fehler.

Ein wesentliches Konstruktionsmerkmal der Items für diesen Zweck bezieht sich auf die Nenner von Brüchen, die addiert werden sollten. Diese sind bei den Summanden der Aufgaben entweder gleich oder ungleich. Bei einer Fragestellung der Studie ist geprüft worden, ob bei den Daten aus beiden Itemformulierungen, offen oder geschlossen, dieser Strukturaspekt jeweils erhalten bleibt. Allein im offenen Format geben die Antworten den Unterschied zwischen gleichen und ungleichen Nennern in den Aufgaben wieder. Im geschlossenen Format läßt sich die gegebene Struktur der Aufgaben nicht rekonstruieren, der Unterschied zwischen den Aufgabentypen verwischt sich bei den geschlossenen Antworten.

Obwohl die Gesamtpunktzahlen sich zwischen den beiden Frageformen nicht unterscheiden, liegt die Anzahl der gefundenen verschiedenen Arten von Fehlern bei der geschlossenen Frageform deutlich über der bei offener Frageform. Wenn die Probanden im geschlossenen Antwortformat die Aufgabe nicht beherrschen, sind sie in ihrem Bemühen, die Aufgabe trotzdem zu meistern, bedeutend weniger konsistent als diejenigen Probanden, die mit einer offenen Frageform konfrontiert gewesen sind. Damit erschwert sich beim geschlossenen Antwortformat die Diagnose, welche mathematische Operationen die Schüler bei der Bearbeitung der Aufgaben nicht richtig beherrschen. Die Antworten aus dem offenen Format lassen den Lösungsweg und die dabei aufgetretenen Schwierigkeiten besser erkennen. Bei vergleichbarer Fähigkeit zur Leistungsmessung liefert das offene Frageformat gegenüber den geschlossenen Fragen noch zusätzliche diagnostische Information, die im dargestellten Fall für eine pädagogische Intervention sehr wertvoll sein kann. Diese vorteilhafte Eigenschaft der offenen Frageform besteht vermutlich nicht nur bei der Leistungs- sondern auch bei der Einstellungsmessung. Anzunehmen ist, daß beispielsweise im Fall des Mißverstehens schon an der gegebenen Antwort einer Frage im offenen Format nicht nur dieser Umstand des Mißverstehens selbst leichter erkannt werden kann, sondern darüber hinaus auch die Art des Mißverstehens. Vielleicht ist bereits anhand der Antworten sogar eine Erklärung für Verzerrungen möglich.

Schwarz betont, daß die Scientific Community in der Regel ausschließlich davon ausgeht, daß der Informationsfluß bei der Anwendung von Fragebogen allein vom Antworter aus gerichtet zu verstehen sei. Er mahnt jedoch: "What we frequently overlook is that our questionnaires are also a source of information that respondents draw on in order to determine their task and to arrive at a useful and informative answer" (1999, S. 103). Wenn dieser Interpretationsvorgang

eines Antworters, aus welchen Gründen auch immer, fehlschlägt, so kann die Wahl des Antwortformats darüber entscheiden, dieses Mißlingen überhaupt sichtbar zu machen. Während im offenen Format eine Fehlpassung von Frage und Antwort unter Umständen sofort augenfällig werden kann („When this is the answer, what was the question?“ . Schwarz, 1999, S. 103), stehen bei geschlossenen Fragen ausschließlich die vom Fragebogenkonstrukteur vorgesehene Reaktionsmöglichkeiten zur Verfügung.

Interviewer oder Fragebogenanwender fordern üblicherweise die Befragten auf, keine der Fragen unbeantwortet zu lassen (forced choice). Das hat vor allem für die Vermeidung fehlender Werte Bedeutung, die sich ansonsten ungünstig auf die Auswertung des Materials auswirken können. Allerdings ist es jedoch angesichts des Drucks der Befragungssituation nicht auszuschließen, daß einige Befragte, die sich zwar nicht in der Lage sehen, jede Frage zu beantworten, trotzdem aber bei diesen Fragen, allein um etwa dem Befrager zu willfahren, eine beliebige Antwortalternative als zutreffend kennzeichnen. Eine durch die Befragten inhaltlich unbegründet vorgenommene Auswahl unter den Antwortkategorien liefert zwar scheinbar die beabsichtigte Beantwortung, Validität dieser Reaktion des Befragten wird durch das gewählte Antwortformat nur vorgegaukelt.

Für einige bekannte Arten von verzerrten Antworten lassen sich Maßnahmen zur Kontrolle entwickeln, wie beispielsweise bei Antwortverhalten im Sinne sozialer Erwünschtheit durch eine „Lügenskala“. Dazu ist allerdings die Voraussicht des ungünstigen Antwortverhaltens für die konkrete Erhebungssituation notwendig. Wenn es jedoch keine Vorerfahrungen mit Anhaltspunkten für den Verdacht einer spezifischen Antwortverzerrung gibt und es dann zu keinem identifizierbaren Muster verzerrter Antworten kommt, wie es beispielsweise bei Ja-Sage-Tendenz auftritt, dann bietet das geschlossene Antwortformat keine Möglichkeit, die Gültigkeit der Antworten zu hinterfragen.