

Kinder und Jugendliche als Befragte: Feldexperimente zum Antwortverhalten Minderjähriger

Fuchs, Marek

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:

Fuchs, M. (2004). Kinder und Jugendliche als Befragte: Feldexperimente zum Antwortverhalten Minderjähriger. *ZUMA Nachrichten*, 28(54), 60-88. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-207664>

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.

KINDER UND JUGENDLICHE ALS BEFRAGTE

Feldexperimente zum Antwortverhalten Minderjähriger

MAREK FUCHS

Die Umfragemethodologie ist hauptsächlich für Studien in Erwachsenenpopulationen entwickelt worden. In den letzten Jahren werden jedoch zunehmend Befragungen durchgeführt, bei denen die Verhaltensweisen und Sichtweisen von Kindern und Jugendlichen direkt von ihnen ermittelt werden. Die Befragung von Eltern oder anderen Stellvertretern verliert demgegenüber an Bedeutung. Nimmt man an, dass die kognitiven Fähigkeiten von Kindern und Jugendlichen im Vergleich zu denen von Erwachsenen noch nicht voll entwickelt sind, dann stellt sich die Frage, ob Kinder und Jugendliche die komplizierten Aufgaben bei der Beantwortung eines Fragebogens bewältigen können.

Der vorliegende Beitrag geht dieser Frage nach, indem die Qualität der von Kindern und Jugendlichen erhaltenen Daten nach Alter und kognitiven Fähigkeiten verglichen wird. Dabei konzentrierten wir uns auf Antwortfehler und schließen von der Größe der Antwortfehler auf die vorgängigen kognitiven Prozesse. Ausgehend von einem Überblick über den Stand der Forschung werden die Ergebnisse von 7 Feldexperimenten zu Antwortfehlern berichtet, die in drei größer angelegte Befragungen von Kindern und Jugendlichen integriert waren ($n=4.206$, $n=5.042$, $n=820$): Experimente zur Fragereihenfolge, zur Antwortreihenfolge, zu Skalen-Effekten und zu den Effekten der numerischen Werte an den Antwortalternativen. Die Ergebnisse zeigen, dass Kinder und Jugendliche Fragebogenfragen von einem kognitiv weniger entwickelten Stand aus beantworten: Sie bearbeiten jede einzelne Fragebogenfrage stärker segmentiert und isoliert, zugleich lassen sie sich bei der Beantwortung der jeweiligen Frage stärker von den darin enthaltenen Informationen leiten.

Survey methodology is predominantly developed for studies in adult populations. In recent years, however, many surveys have addressed samples of children and juveniles. Today, data on children's behavior, attitudes and beliefs are usually collected directly from them. Proxy-reporting from parents or other adults taking care of the children is no longer seen as an adequate and appropriate means of data collection. At the same time, given that their cognitive skills may still be developing, one may wonder whether

children can adequately perform the difficult tasks associated with responding to a questionnaire. As a result the quality of data obtained from children might be worse than that collected from adult populations.

In this paper we assess the quality of data obtained from children and juveniles. We focus on response error and compute the sizes of response error for specific age groups and groups with different levels of cognitive functioning. Against the background of an extensive literature review, we analyze field experiments that were part of three large-scale surveys among children and juveniles aged 10 to 21 ($n=4.206$, $n=5.042$, $n=820$). The results of 7 experiments on response order, question order, scale effects and the effects of the numeric values associated to the response categories are reported. The results indicate that younger and less well-educated children answer survey questions from a cognitively less sophisticated perspective than do adults. Children and juveniles process survey questions in a more segmented fashion and view each as separate entities. In addition, in answering questions they are influenced to a greater degree than adults by the information provided in the question text and by the response categories of each given question.

1. Einleitung und Fragestellung

Die Theorie des Frage-Antwort-Prozesses ist relativ gut entwickelt, ebenso die Regeln für die Konstruktion von „guten“ Fragebögen. Als Ergebnis dieser methodologischen Bemühungen kann man heute auf eine Reihe von Überblicksdarstellungen zur Datenqualität (Groves 1989; Lyberg et al. 1997) und zur Fragebogenkonstruktion (Schuman/Presser 1981; Sudman et al. 1996; Tourangeau et al. 2000) zurückgreifen. Diese Darstellungen beziehen sich im Wesentlichen auf Umfragen unter Erwachsenen; in den letzten Jahrzehnten wurde jedoch eine Reihe von Studien mit Kindern und Jugendlichen durchgeführt, bei denen deren Verhaltensweisen, deren Einstellungen und sonstige Informationen von ihnen selbst erhoben, und nicht auf Eltern oder anderen Stellvertretern (Proxy) zurückgegriffen wurde (in Deutschland z.B. das DJI-Kinderpanel, aber auch Studien zur Devianz von Kindern und Jugendlichen oder kommerzielle Untersuchungen). In der Literatur (Holiday/Turner-Henson 1989; Scott 1997; Borgers et al. 2000) wird in diesem Zusammenhang darauf hingewiesen, dass Umfragen unter Kindern und Jugendlichen deren kognitiven und sozialen Entwicklungsstand zu berücksichtigen haben.¹ Denn die begrenzten kognitiven Ressourcen der kindlichen Befragten könnten u. U. zu einem eingeschränkten Fragever-

1 Eine sehr nette Verulkung der Methodenforschung zur Befragung von kleinen Kindern findet sich bei Miller 1976.

ständnis und zu anderen Beeinträchtigungen der Datenqualität führen, weil die jungen Befragten nicht oder nur eingeschränkt über die Voraussetzungen zur Beantwortung eines Fragebogens verfügen.

Ausgehend von der Literatur zur Psychologie des Umfrageinterviews (z.B. Sudman et al. 1996) stellt sich der Frage-Antwort-Prozess als komplexe Interaktion des Befragten mit dem Fragebogen und anderen Elementen der Umfrage dar. In dieser Perspektive gliedert sich die Aufgabe des Befragten bei der Beantwortung eines einzelnen Items in verschiedenen Phasen: Zunächst muss der Frageinhalt verstanden werden bevor nach relevanten Informationen zu ihrer Beantwortung gesucht werden kann. Darauf aufbauend wird die eigentliche Antwort generiert, bevor diese hinsichtlich des angebotenen Antwortschemas formatiert und schließlich unter Umständen editiert wird. Bei der Abarbeitung dieser Phasen des Frage-Antwort-Prozesses berücksichtigt der Befragte die Frageformulierung und die dargebotenen Antwortkategorien der aktuellen Frage und ebenso die vorausgehenden – und zum Teil auch die nachfolgenden – Fragen als signifikanten Kontext (für eine umfassende Übersicht vgl. auch Tourangeau et al. 2000). In Anbetracht der Komplexität dieses Vorgangs stellt sich die Frage „whether children can adequately perform these difficult tasks“ (de Leeuw/Otter 1995: 251) – denn Kinder verfügen im Vergleich zu Erwachsenen über weniger entwickelte kognitive Fähigkeiten.

Die kognitive Entwicklung von Kindern und Jugendlichen wird in der Regel als Stufenfolge konzeptualisiert (vgl. u. a. Dockrell et al. 2000): Folgt man den theoretischen Ansätzen von Piaget (1948) oder Kohlberg (1976), so ist die menschliche Entwicklung nicht als ein diskreter Schritt von Kindheit zum Erwachsenenstatus zu verstehen. Vielmehr verläuft die Entwicklung von Kindern und Jugendlichen als kontinuierlicher Prozess über einen längeren Zeitraum, in dem nacheinander verschiedene konsekutive Stufen der kognitiven und sozialen Entwicklung erreicht werden. Daher können Kinder und Jugendliche nicht als homogene Gruppe verstanden werden, sondern es muss zwischen verschiedenen Stufen der kognitiven Kapazität differenziert werden. Ausgehend von diesen Theorien wird in der Literatur zur Umfrageforschung mit Minderjährigen die Vermutung vertreten, dass Kinder ab einem Alter von etwa 10 Jahren in der Lage sind, standardisierte Befragungen zu bearbeiten (vgl. etwa Scott 1997 und Borgers et al. 2000 für Überblicksdarstellungen; vgl. Benson/Hocevar 1985 für Aussagen zu Grundschulkindern). Es wird angenommen, dass Kinder dieses Alters bereits eine spezifische Stufe der formalen Operation (Piaget 1948) oder – in Kohlbergs (1976) Terminologie – die erste Stufe des konventionellen Levels der moralischen Entwicklung erreicht haben. Daher sind sie in der Lage und willens, abstrakte Inhalte zu verarbeiten und Regeln zu folgen, die unabhängig von spezifischen Subjekten in konkreten Interaktionssituationen Geltung beanspruchen. Es ist jedoch offensichtlich, dass 10-jährige Kinder Fragebogenfragen vor dem Hintergrund weniger

entwickelter kognitiver Fähigkeiten betrachten und bearbeiten als 17-jährige Jugendliche oder Erwachsene, und dass sich daher ihr Antwortverhalten in standardisierten Befragungen unterscheidet. Es stellt sich damit die Frage, wie sich Kinder, Jugendliche und junge Erwachsene bei der Bearbeitung eines standardisierten Fragebogens unterscheiden.

Bisher liegen jedoch nur wenige empirische Analysen zum Antwortverhalten von Kindern und jüngeren Jugendlichen in standardisierten Befragungen vor; außerdem handelt es sich dabei im Wesentlichen sekundäranalytische Studien auf der Basis existierender Datensätze. Allgemein wird jedoch davon ausgegangen, dass die begrenzte kognitive Kapazität jüngerer Befragter die Datenqualität negativ beeinflusst: Basierend auf einem dreiwöchigen Panel zur politischen Sozialisation kann Villancourt (1973) zeigen, dass die Antwortstabilität von Kindern – wie bei den Antworten von Erwachsenen – erheblich variiert. Allerdings wurde kein Vergleich mit Jugendlichen oder Erwachsenen durchgeführt. Marsh (1986) stellt bei einem Vergleich verschiedener Altersgruppen fest, dass „younger children and children with poorer reading skills are less able to response appropriately to negative items and that this effect biases the interpretation of their responses“ (1986: 37; vgl. auch Hershey/Hill 1976).

Amato und Ochiltree (1987) zeigen, dass die Datenqualität einer Gruppe von 8- und 9-Jährigen signifikant niedriger ist, als die einer Vergleichsgruppe von 15- und 16-Jährigen. In einer Studie von Borgers und Kollegen (2000) erwies sich die Anzahl der absolvierten Schuljahre als aussagekräftiger Prädiktor für die interne Konsistenz von Multi-Item-Skalen. Dies unterstützt die Hypothese, wonach die Datenqualität mit zunehmender kognitiver Kapazität ansteigt. Hinzu kommt, dass schlechte Lesekompetenzen einen negativen Einfluss auf die Datenqualität haben (Borgers et al. 2000). Weiterhin konnte gezeigt werden, dass mehrdimensionale Antwortskalen für jüngere Kinder mit größeren Schwierigkeiten verbunden waren und einen höheren Item-Nonresponse produzierten als bei älteren Kindern. Folgt man Borgers und Hox (2001) ist dies auf deren kognitives Entwicklungsstadium zurückzuführen.

Unterstützend können hier die Ergebnisse von de Leeuw und Otter (1995) herangezogen werden, die eine klare Interaktion zwischen dem Alter und dem Effekt unklarer Fragen nachweisen konnten. Danach war der negative Effekt mehrdeutiger Fragen auf die Datenqualität bei jüngeren Kindern stärker ausgeprägt als bei älteren Kindern und Jugendlichen, die mit der Ambiguität besser umgehen konnten. Neben dem Alterseffekt fanden Borgers und Hox (2000) ebenfalls einen Einfluss des Leistungsstandes in der Schule. Fasst man diese vorliegenden Ergebnisse zusammen, so stützen sie die Hypothese, wonach das geringe Alter von Kindern und jüngeren Jugendlichen und zudem eine geringe schulische Leistungsfähigkeit (als Indikatoren für das kognitive Entwicklungsstadium) einen negati-

ven Einfluss auf die Bearbeitung von Fragebogenfragen und damit auf die Datenqualität haben.

Diese These erfährt durch Erkenntnisse zum Antwortverhalten von Erwachsenen Unterstützung: Folgt man Krosnick (1992), dann sind umfangreiche Gedächtnisprozesse und kognitive Aktivitäten für Befragte mit geringer kognitiver Kapazität weniger angenehm, und sie führen diese Aufgaben weniger gerne aus. Ohne diese Aktivitäten wird jedoch ein spezifisches Antwortverhalten wahrscheinlich, das Krosnick (1992) als „Satisficing“ bezeichnet: Anstatt den Frage-Antwort-Prozess vollständig und gründlich zu durchlaufen, unternehmen die dieses Verhalten zeigenden Befragten weniger Anstrengungen bei der Dechiffrierung einer Fragebogenfrage und bei der Suche nach einer Antwort. Entsprechend wird nur „der Form genüge getan“ und es wird z.B. die erste akzeptable Antwortalternative gewählt oder sonst ein „leichter“ Weg durch die Befragung gesucht.

Jedoch fehlt bisher eine experimentelle Überprüfung des Zusammenhangs von kognitiven Fähigkeiten und der Bearbeitung des Frage-Antwort-Prozesses durch Kinder – dazu versucht der vorliegende Beitrag beizutragen: In sieben größer angelegten Feldexperimenten wurden das Alter und das schulische Leistungsniveau als Indikatoren für die kognitive Kapazität benutzt (vgl. hierzu die Ausführungen im Abschnitt über die Methoden), und zu Antworteffekten, an denen sich die kognitiven Prozesse bei der Beantwortung einer Fragebogenfrage ablesen lassen, in Beziehung gesetzt. Dazu wird in einem ersten Schritt überprüft, ob diese Antworteffekte auch für Kinder und Jugendliche reproduziert werden können, bevor in einem zweiten Schritt der Einfluss von Alter und schulischer Leistungsfähigkeit auf die Größe dieser Effekte untersucht wird.

Ausgehend von der Vermutung, dass die kognitiven Fähigkeiten von Kindern weniger entwickelt sind als die von Jugendlichen bzw. Erwachsenen, lassen sich verschiedene Hypothesen über die Besonderheiten der Antworteffekte bei Kindern und Jugendlichen aufstellen, die im vorliegenden Beitrag getestet werden sollen. Diese Hypothesen folgen der generellen Annahme, dass Kinder und jüngere Jugendliche die einzelnen Fragebogenfragen stärker segmentiert und aus dem Fluss des Fragebogens herausgelöst interpretieren, weil ihnen die kognitiven Fähigkeiten fehlen, sich den Fragebogen bzw. einzelne Frageblöcke als thematische Einheit zu vergegenwärtigen und im Gedächtnis zu behalten. Zugleich lassen sie sich bei der Beantwortung jeder einzelnen Frage stärker von deren Merkmalen leiten, weil sie bei der Bewältigung des kognitiv anspruchsvollen Frage-Antwort-Prozesses auf die Unterstützung und Hilfe durch die in der Frage und in den Antwortkategorien enthaltenen Informationen angewiesen sind. Im Einzelnen werden folgende Hypothesen überprüft:

Fragereihenfolge: Befragte, die eine geringere kognitive Funktionalität aufweisen, können bei der Bearbeitung einer Frage vorausgehende Fragen und Informationen in geringerem Umfang berücksichtigen und daher sind bei ihnen Kontexteffekte – als Indikator für den Umfang der berücksichtigten vorausgehenden Information – weniger ausgeprägt. Entsprechend vermuten wir bei jüngeren Befragten und bei Befragten mit geringerer schulischer Leistungsfähigkeit kleinere Fragereihenfolgen-Effekte als bei älteren Befragten und Befragten mit besserer schulischer Leistungsfähigkeit.

Skalen-Effekte: Neben den vorausgehenden Fragen nutzen die Befragten auch Merkmale der aktuellen Frage und der vorgegebenen Antwortmöglichkeiten, um den Frageinhalt zu verstehen und um ihre Antworten zu generieren. So wird bei geschlossenen Häufigkeitsfragen neben dem Fragetext von den Befragten auch das vorgegebene Antwortkontinuum herangezogen, um eine Vorstellung von der Größenordnung der gesuchten Antwort zu bekommen – vor allem, wenn die gesuchte Antwort geschätzt werden muss und nicht aus der Erinnerung abgerufen werden kann. Entsprechend unserer generellen Annahme werden sich jüngere Befragte und solche mit schlechteren schulischen Leistungen bei der Beantwortung von geschlossenen Häufigkeitsfragen stärker von den vorgegebenen Antwortkategorien leiten lassen. Als Ergebnis werden wir bei ihnen größere Skalen-Effekte nachweisen können.

Numerische Werte: Neben den verbalisierten Antwortkategorien lassen sich die Befragten auch von den numerischen Werten an Antwortalternativen leiten. Von den einleitend diskutierten theoretischen Überlegungen ausgehend, lässt sich vermuten, dass sich Befragte mit geringer entwickelten kognitiven Fähigkeiten bei der Generierung des Frageverständnisses stärker an den Zahlenwerten, die den Antwortkategorien beigelegt sind, orientieren werden. Entsprechend sollte der Effekt unterschiedlicher numerischer Symbole bei jüngeren Befragten und solchen mit schlechteren schulischen Leistungen größer ausfallen.

Antwortreihenfolge: Weiter vermuten wir, dass die begrenzte kognitive Kapazität zu größeren Antwortreihenfolgen-Effekten führt. Kinder und Jugendliche mit einer geringer entwickelten kognitiven Kapazität sind weniger gut in der Lage, jeder Antwortkategorie die Aufmerksamkeit zuzuwenden und daher werden sie sich bei der Auswahl einer Antwortkategorie eher von deren sequenzieller Position leiten lassen. Dies führt bei jüngeren Befragten und solchen mit schlechteren schulischen Leistungen zu größeren Antwortreihenfolgen-Effekten (Recency- und Primacy-Effekte).

2. Methode

Diese Hypothesen werden anhand der Ergebnisse von Feldexperimenten überprüft, die auf drei größeren, selbst-administrierten Umfragen basieren,² die in den Jahren 1999 bis 2001 mit Kindern, Jugendlichen und jungen Erwachsenen im Alter von 10 bis 21 Jahren zu verschiedenen inhaltlichen Themen durchgeführt wurden ($N_1 = 4.206$; $N_2 = 5.042$; $N_3 = 820$):

Die *erste* Studie wurde im Frühjahr 1999 als zweite Welle einer laufenden Längsschnittuntersuchung zur „Gewalt in Schulen“ in Bayern durchgeführt (Fuchs et al. 2001). Eine disproportional geschichtete Klumpenstichprobe von bayerischen Schülern allgemein bildender und beruflicher Schulen wurde untersucht, um repräsentative Aussagen über die Gewalttätigkeit und die Opfererfahrungen der Schüler in Bayern machen zu können. Insgesamt wurden 4.206 Kinder und Jugendliche im Alter zwischen 10 und 21 Jahren (einige wenige waren älter) im Klassenzimmer mit einem selbst-administrierten Fragebogen befragt.

Die *zweite* Studie wurde im Frühjahr 2001 ebenfalls als selbst-administrierte Befragung zum Thema „Rechtsextremismus unter Jugendlichen“ durchgeführt (Fuchs et al. 2003). 5.042 Befragte im Alter zwischen 13 und 21 Jahren (einige wenige waren älter) bearbeiteten einen Fragebogen im Klassenzimmer-Setting. Auch hier haben wir eine disproportional geschichtete Klumpenstichprobe gezogen, um eine repräsentative Schülerpopulation der allgemeinbildenden und beruflichen Schulen in Bayern realisieren zu können.

Die *dritte* Studie beschäftigte sich mit Konflikten zwischen den Generationen. Die Studie wurde im Frühjahr 2000 im Bundesland Bayern als schriftlich-postalische Befragung durchgeführt, die auf einer Random-Registerstichprobe von 820 Befragten basierte, die nach Alter und Gemeindegröße geschichtet war. Die Besonderheit dieser Studie besteht darin, dass Jugendliche und junge Erwachsene zwischen 15 und 20 Jahren mit Erwachsenen sowie mit alten und hochbetagten Befragten verglichen werden können.

In diese Studien wurden Experimente zur Fragereihenfolge, zu Skalen-Effekten, zur Antwortreihenfolge und zu den numerischen Werten an Antwortvorgaben integriert. Die Befragten wurden zufällig zu zwei bzw. vier Fragebogen-Splits zugewiesen. Jede dieser Gruppen erhielt einen ansonsten identischen Fragebogen, der sich nur hinsichtlich der experimentell modifizierten Fragen unterschied. Durch den Vergleich der Antwortverteilungen der Splits können Antworteffekte identifiziert und berechnet werden. Aufgrund der beträchtlichen Fallzahlen besteht weiter die Möglichkeit, die ermittelten Antworteffekte

2 Zu Mode-Effekten bei Befragungen von Jugendlichen vgl. Naplaval/Oberwittler 2002.

für verschiedene Befragtengruppen mit unterschiedlichem kognitivem Niveau separat zu berechnen. Ein Vergleich der Effektgrößen für diese verschiedenen Teilpopulationen ermöglicht uns dann den Test der Hypothesen.

Zur Bildung der verglichenen Teilpopulationen wurde zunächst das Alter herangezogen. Unter Berücksichtigung der Fallzahlen in den einzelnen Jahrgängen haben wir die Befragten zu Altersgruppen zusammengefasst, die aus 2 bis 4 Jahren bestehen. Nur bei der altersmäßigen Kategorisierung der Befragten in Studie 3 haben wir – auf Grund der geringen Fallzahlen – größere Altersgruppen bilden müssen. Wegen der unterschiedlichen Altersspannen lassen sich die Altersgruppen der drei Studien nicht direkt vergleichbar gestalten – die Fallzahlen der einzelnen Alterskategorien wären sonst sehr ungleich und in einzelnen Fällen zu gering, um die Effekte sichtbar zu machen.

Das Niveau der kognitiven Kapazität haben wir weiter durch die schulischen Leistungen/Bildungsniveau gemessen (vgl. zu dieser Operationalisierung Knäuper 1999): In den Studien 1 und 2 basiert die Kategorisierung auf einer zusammenfassenden subjektiven Einschätzung der Zufriedenheit der Lehrer mit den schulischen Leistungen durch die Befragten;³ in Studie 3 wurde auf den allgemein bildenden Schulabschluss (den erreichten bzw. den Angestrebten) zurückgegriffen. Diese relativ indirekte Messung der kognitiven Kapazität hängt zum einen mit den hohen Ressourcenansprüchen einer direkten, umfassenden Messung zusammen (Platz im Fragebogen und Zeit im Interview), und zum anderen mit der Tatsache, dass das Konzept der kognitiven Kapazität in Bezug auf den Frage-Antwort-Prozess bisher nur ansatzweise theoretisch ausgearbeitet ist. Die Selbsteinstufung der schulischen Leistungen bzw. das allgemeine Bildungsniveau wurden zu drei Kategorien rekodiert („gut“ – bei den schulischen Leistungen – bzw. „hoch“ – beim Bildungsniveau –, „mittel“, „niedrig“)⁴ und mit Blick auf die Fallzahlen in den einzelnen Zellen für die Zwecke der hier vorgenommenen Analysen zu „gut“ bzw. „hoch“ einerseits und „mittel/niedrig“ andererseits zusammengefasst.

3 Dabei wurde „sehr zufrieden“ und „zufrieden“ als Indikator für „gute“ schulische Leistungen aufgefasst. Die übrigen Ausprägungen („teils/teils“, „weniger zufrieden“ und „gar nicht zufrieden“) wurden zu „mittleren/niedrigen“ schulischen Leistungen zusammengefasst.

4 Beim Bildungsniveau wurde die (fachgebundene und die allgemeine) Hochschulreife und der (Fach-)Hochschulabschluss als hohes Bildungsniveau codiert; alle übrigen Kategorien wurden zu „mittel/niedrig“ zusammengefasst (kein Abschluss, Hauptschulabschluss, Mittlere Reife).

3. Ergebnisse

Im Folgenden konzentrieren wir uns zunächst auf Antwortreihenfolgen-Effekte (3.1) bevor wir Skalen-Effekte (3.2) und die Effekte der numerischen Werte, die den Antwortkategorien zugeordnet sind, berichten (3.3). Im abschließenden Teilkapitel werden wir uns den Ergebnissen von zwei Experimenten zur Fragereihenfolge zuwenden (3.4.).

3.1 Antwortreihenfolge-Effekte

Experiment 1

Als Teil von Studie 2 wurden die Befragten gebeten, aus einer Liste von vorgegebenen Jugendkulturen diejenige auszuwählen, der sie sich am nächsten fühlen. In einem der beiden randomisierten Fragebogensplitts waren die zehn Antwortkategorien in vollständig gedrehter Reihenfolge abgedruckt (vgl. Abbildung 1).

Abbildung 1: Frageformulierung in Experiment 1 zur Antwortreihenfolge, Originalversion (links) und gedrehte Version (rechts)

26. Stell dir vor, es kommt ein neuer Mitschüler in deine Klasse und sieht dich das erste Mal. Was würde der denken welcher Gruppierung du nahe stehst? (bitte nur eine Antwort ankreuzen)

- Rapper (Hip-Hop)
- Techno
- Skinheads
- Autonome
- Alternative
- Gothics
- Metaller
- Rockabilly, Rock'n'Roll
- Hooligans
- Punks

zu einer anderen Gruppe: _____

26. Stell dir vor, es kommt ein neuer Mitschüler in deine Klasse und sieht dich das erste Mal. Was würde der denken welcher Gruppierung du nahe stehst? (bitte nur eine Antwort ankreuzen)

- Punks
- Hooligans
- Rockabilly, Rock'n'Roll
- Metaller
- Gothics
- Alternative
- Autonome
- Skinheads
- Techno
- Rapper (Hip-Hop)

zu einer anderen Gruppe: _____

Von Krosinck und Alwin (1987) ausgehend erwarteten wir einen Primacy-Effekt, denn in selbst-administrierten Befragungen erhalten Antwortmöglichkeiten, die am Anfang einer langen Liste von Kategorien präsentiert werden, mehr Aufmerksamkeit und sie setzen einen interpretativen Rahmen für die Wahrnehmung der nachfolgenden Antworten (im Sinn eines Standards für den Vergleich). Hinzu kommt, dass „a satisfying criterium implies that the earlier in the list an acceptable answer appears the more popular it will be. If respondents chose an earlier option they avoid thinking about the others” (Tourangeau et al. 2000: 250).

Tabelle 1: Größe des Antwortreihenfolgen-Effekts (Recency-Effekt, Prozentpunkte) nach Alter und schulischen Leistungen

Altersgruppe	Schulische Leistungen		Alle
	Gut	Mittel/Schlecht	
13 bis 15 Jahre	19 *	10 *	11 ***
16 bis 17 Jahre	10	2	4
18 bis 21 Jahre	-1	0	1
Alle	7	5	6 ***

Studie 2, Frage 26. Angegeben ist die Prozentpunktdifferenz = Anteil der Nennungen bei Positionierung an letzter Stelle minus Anteil der Nennungen bei Positionierung am Anfang; ohne die „anderen Gruppen“, die insgesamt nur selten genannt wurden (7,7%).

* $p < 0,05$; *** $p < 0,001$ für die jeweiligen Effekte; Test auf Differenzen zwischen den Effekten: Alter $p < 0,07$, schulische Leistungen nicht signifikant.

Da es sich um die Differenz von Prozentsatzdifferenzen handelt, sind Standard-Tests für die Unterschiede zwischen den Antworteffekten nicht verfügbar. Hier wird ein χ^2 -Anunabhängigkeitstest auf die Größe der Prozentsatzdifferenzen unter Berücksichtigung der Fallzahl in der entsprechenden Kategorie der Randverteilung berechnet.

Entgegen dieser Hypothese und den in der Literatur dokumentierten Ergebnissen (z.B. Hippler/Schwarz 1992) beobachten wir jedoch keinen Primacy-Effekt, sondern einen Recency-Effekt in unseren Daten: Die Antwortkategorien, die am Ende der Liste stehen, werden unabhängig vom Inhalt häufiger genannt (6 Prozentpunkte; $p < 0,001$).⁵ Nachdem es in unserer Studie jedoch vordringlich darum geht, die Größe des Antwortreihenfolgen-Effektes in Abhängigkeit vom Alter und von den schulischen Leistungen zu untersuchen, vernachlässigen wir möglichen Ursachen für das Auftreten dieses Recency-Effekts und analysieren stattdessen die Effektgröße in Abhängigkeit von diesen Variablen.

Ausgehend von den theoretischen Überlegungen, die wir im ersten Kapitel zusammengefasst haben, erwarten wir, dass ältere Befragte und Befragte mit besserer schulischer Leistung eher in der Lage sind alle Antwortkategorien zur gleichen Zeit im Gedächtnis zu behalten und daher zu einem größeren Anteil jeder Antwortkategorie die gleiche Aufmerksamkeit zuzuwenden. Dies sollte sich in einem weniger ausgeprägten Antwortreihenfolgen-Effekt ausdrücken. Wie erwartet variiert die Effektgröße mit dem Alter: Jüngere

5 Wir vermuten, dass der Recency-Effekt vor allem als Folge der Tatsache auftritt, dass die angebotenen Jugendkulturen für einen beträchtlichen Teil der Befragten wenig plausibel erscheinen. Sudman und Kollegen haben betont (1996, S. 141), dass „a given implausible item should elicit more disagreeing thoughts and be less likely to be endorsed when presented early rather than late in the list“.

Befragte weisen tendenziell größere Effekte als ältere Jugendliche und volljährige Befragte auf ($p < 0,07$). Die schulische Leistungsfähigkeit hat jedoch keinen signifikanten Einfluss auf die Größe der Antwort-Reihenfolgen-Effekte.

Experiment 2

Im zweiten Experiment zu Antwortreihenfolgen-Effekten (Studie 1) wurden die Befragten ebenfalls mit einer langen Liste von Antwortkategorien konfrontiert, in der infrastrukturelle Einrichtungen des Wohnquartiers angegeben waren (Abbildung 2).

Abbildung 2: Fragebogenversionen in Experiment 2 zur Antwortreihenfolgen, Originalversion (rechts) und gedrehte Version (links; die übrigen Versionen sind nicht abgedruckt)

- | | |
|--|--|
| <p>31. Welche der folgenden Einrichtungen gibt es in deinem Wohnviertel oder deinem Dorf? (Du kannst mehrere Antworten ankreuzen.)</p> <ul style="list-style-type: none"> <input type="checkbox"/> Supermarkt <input type="checkbox"/> Kinderspielplatz <input type="checkbox"/> Kindergarten <input type="checkbox"/> Museum <input type="checkbox"/> Theater <input type="checkbox"/> Bücherei <input type="checkbox"/> Busverbindungen <input type="checkbox"/> Schule <input type="checkbox"/> Hallenbad <input type="checkbox"/> Jugendzentrum <input type="checkbox"/> Diskothek <input type="checkbox"/> Kino | <p>31. Welche der folgenden Einrichtungen gibt es in deinem Wohnviertel oder deinem Dorf? (Du kannst mehrere Antworten ankreuzen.)</p> <ul style="list-style-type: none"> <input type="checkbox"/> Kino <input type="checkbox"/> Diskothek <input type="checkbox"/> Jugendzentrum <input type="checkbox"/> Hallenbad <input type="checkbox"/> Schule <input type="checkbox"/> Busverbindungen <input type="checkbox"/> Bücherei <input type="checkbox"/> Theater <input type="checkbox"/> Museum <input type="checkbox"/> Kindergarten <input type="checkbox"/> Kinderspielplatz <input type="checkbox"/> Supermarkt |
|--|--|

Die Befragten sollten aus dieser Liste diejenigen Ausstattungselemente auswählen, die in ihrer Wohnumgebung vorhanden sind. Die 12 Antwortkategorien wurden in vier verschiedenen Reihenfolgen präsentiert: Neben der Original-Version gab es eine vollständig gedrehte Version und zwei weitere Versionen, bei denen die Antwortkategorien vom Anfang und vom Ende der Liste an unterschiedlichen Positionen in der Mitte erschienen.

Wie erwartet (Krosnick/Alwin 1987; Schwarz et al. 1992) zeigen die Daten einen moderaten Primacy-Effekt. Antwortkategorien, die am Beginn der Liste erscheinen, erhalten unabhängig vom Inhalt der Antwortkategorie höhere Zustimmungswerte verglichen mit einer Positionierung in der Mitte oder am Ende der Liste. In Tabelle 2 werden die Ergebnisse für die Antwortkategorie „Kino“ (= es gibt ein Kino im Wohnquartier der Befragten) dargestellt. Wird „Kino“ am Anfang der Liste der Antwortkategorien präsentiert, entscheiden sich signifikant mehr Befragte dafür (5 Prozentpunkte, $p < 0,001$).

Tabelle 2: Größe des Antwortreihenfolgen-Effektes (Primacy-Effekt, Prozentpunkte) nach Alter und schulischen Leistungen

Altersgruppe	Schulische Leistungen		Alle
	Gut	Mittel/Schlecht	
10 bis 13 Jahre	7 *	11 *	10 ***
14 bis 17 Jahre	1	5	3
18 bis 21 Jahre	5	2	4
Alle	4 *	6 **	5 ***

Studie 1, Frage 31, Kategorie "Kino". Angegeben ist die Prozentsatzdifferenz zwischen der Originalversion und den drei übrigen Version; also Anteil der Nennungen bei Positionierung am Anfang minus Anteil der Nennungen bei Positionierung in der Mitte oder am Ende.

* < 0,05; ** p < 0,01, *** p < 0,001 für die jeweiligen Effekte; Test auf Unterschiede zwischen den Effekten: Alter p < 0,01, schulische Leistungen p < 0,05; vgl. die Anmerkung zu Tabelle 1.

Die Größe dieses Antwortreihenfolgen-Effektes korreliert wie erwartet mit dem Alter der Befragten: Jüngere Befragte zwischen 10 und 13 Jahren zeigen größere Effekte (10 Prozentpunkte) im Vergleich zu älteren Befragten ($p < 0,01$). Außerdem ist auch die schulische Leistungsfähigkeit korreliert: Für Schüler mit besseren schulischen Leistungen errechnet sich ein kleinerer Antwortreihenfolgen-Effekt (4 Prozentpunkte) als bei den Befragten mit durchschnittlichen oder schlechten schulischen Leistungen (6 Prozentpunkte; $p < 0,05$).

Insgesamt unterstützen die Befunde aus den Experimenten 1 und 2 die Behauptung, dass sich jüngere Befragte stärker von der Antwortreihenfolge leiten lassen, also aufgrund der geringeren kognitiven Verarbeitungskapazitäten den einzelnen Antwortmöglichkeiten unterschiedlich viel Aufmerksamkeit zuwenden. Der entsprechende Einfluss schlechter schulischer Leistungen lässt sich nur in Experiment 2 nachweisen.

3.2 Skalen-Effekte

Das von uns benutzte Experiment zu den Skalen-Effekten basiert auf einer klassischen Studie von Schwarz und Mitarbeitern (1985), in dem die Befragten gebeten wurden, die Anzahl der Stunden zu schätzen, die sie pro Tag durchschnittlich fernsehen. Die Befragten wurden dabei zufällig zwei verschiedenen Antwortschemata zugewiesen, die entweder von „bis zu einer ½ Stunde“ bis „mehr als 2½ Stunden“ reichten oder aber von „bis zu 2 ½ Stunden“ bis „mehr als 4 ½ Stunden“. In der von Schwarz und Mitarbeitern (1985) durchgeführten Studie berichteten Befragte, die mit der höherreichenden Antwortskala konfrontiert wurden, einen signifikant längeren durchschnittlichen TV-Konsum im Ver-

gleich zu den mit der niedrigeren Werte umfassenden Antwortskala konfrontierten Befragten.

Experiment 3

In dem in Studie 2 integrierten Experiment 3 haben wir diese klassische Studie mit Kindern und Jugendlichen repliziert. Die befragten Kinder und Jugendlichen waren gebeten, ihren täglichen TV-Konsum auf einer 6-stufigen Skala zu berichten, die analog zu den Kategorien von Schwarz und Mitarbeitern (1985) in zwei verschiedenen Varianten dargeboten waren. Entsprechend können wir voraussagen, dass Befragte, die mit den höheren Antwortkategorien konfrontiert werden, eine längere tägliche Fernsehdauer berichten werden, als Kinder und Jugendliche, die die Frage mit den niedrigen Antwortkategorien beantworten (= Skalen-Effekt).

Abbildung 3: Fragebogenversionen im Experiment 3 zum Skalen-Effekt, „täglicher TV-Konsum“

24. Wie viele Stunden am Tag schaust du ungefähr fern?

- bis 2¼ Stunde
- bis 3 Stunde
- bis 3½ Stunden
- bis 4 Stunden
- bis 4¼ Stunden
- über 4½ Stunden

24. Wie viele Stunden am Tag schaust du ungefähr fern?

- bis ½ Stunde
- bis 1 Stunde
- bis 1½ Stunden
- bis 2 Stunden
- bis 2½ Stunden
- über 2½ Stunden

Insgesamt beobachten wir – im Einklang mit den Befunden von Schwarz und Mitarbeitern (1985) – einen beträchtlichen Skalen-Effekt: Der Anteil der Befragten, die eine Antwortkategorie ausgewählt haben, die eine längere als 2½-stündige tägliche Fernsehdauer repräsentiert, ist um 25 Prozentpunkte höher ($p < 0,001$), wenn die Frage die hohen Antwortalternativen enthält.

Ausgehend von den theoretischen Überlegungen im Einleitungskapitel erwarten wir weiter, dass jüngere Befragte und Befragte mit schlechteren Schulleistungen sich stärker an den Antwortkategorien orientieren als ältere Jugendliche und Befragte mit besseren Schulleistungen: Weil ihre kognitiven Fähigkeiten sich erst noch entwickeln, sind erstere weniger gut in der Lage, die gewünschte Schätzung des täglichen TV-Konsums ohne Rückgriff auf das Antwortkontinuum vorzunehmen, als Befragte mit besser entwickelten kognitiven Fähigkeiten – entsprechend werden wir bei ihnen einen größeren Skalen-Effekt beobachten.

Tabelle 3: Größe des Skalen-Effekts (Prozentpunkte) nach Alter und schulischen Leistungen

Altersgruppe	Schulische Leistungen		Alle
	Gut	Mittel/Schlecht	
13 bis 15 Jahre	25 ***	32 ***	30 ***
16 bis 17 Jahre	19 ***	28 ***	25 ***
18 bis 21 Jahre	22 ***	17 ***	19 ***
Alle	22 ***	27 ***	25 ***

Studie 2, Frage 24. Angegeben ist die Prozentsatzdifferenz für den Anteil der Nennungen über 2½ Stunden Fernsehdauer in der Version mit den hohen Antwortkategorien minus des Anteils der Nennungen bei der Version mit den niedrigen Antwortkategorien.

*** $p < 0,001$ für den jeweiligen Effekt; Test auf Unterschiede der Effekte: Alter $p < 0,01$, schulische Leistungen $p < 0,05$; vgl. die Anmerkung zu Tabelle 1.

Entsprechend unserer Hypothese zeigen jüngere Befragte größere Skalen-Effekte als Ältere. Während der Skalen-Effekt bei den volljährigen Personen 19 Prozentpunkte erreicht, kommen Befragte zwischen 16 und 17 Jahren auf 25 Prozentpunkte und in der Weiterführung dieses Trends berechnen wir in der jüngsten Altersgruppe der bis 15-Jährigen einen Skalen-Effekt von 30 Prozentpunkten ($p < 0,01$). Außerdem zeigt sich, dass Personen mit schlechteren schulischen Leistungen größere Effekte aufweisen (27 Prozentpunkte) als solche mit besseren Leistungen (22 Prozentpunkte; $p < 0,05$). Diese Ergebnisse unterstützen die Hypothese wonach Befragte mit besseren kognitiven Fähigkeiten eher in der Lage sind, Antworten unabhängig vom angebotenen Antwortschema zu generieren. Befragte mit weniger entwickelten kognitiven Fähigkeit orientieren sich entsprechend stärker am Hinweischarakter der Antwortkategorien. Eine weitere Inspektion der Ergebnisse zeigt weiter, dass es vor allem die jüngeren, weniger erfolgreichen Schüler sind, die den größten Skalen-Effekt aufweisen (32 Prozentpunkte).

Experiment 4

In einem zweiten Experiment zu Skalen-Effekten haben wir das gleiche klassische Design von Schwarz und Mitarbeitern (1985) benutzt, wie im vorangegangenen Experiment. Die Befragten in Studie 3 wurden wiederum gebeten, ihren täglichen Fernsehkonsum auf einer 6-stufigen Skala zu schätzen. Dabei haben wir die beiden experimentellen Versionen des Antwortkontinuums verwendet, das bereits oben ausführlich beschrieben wurde (vgl. Experiment 3 und Abbildung 3 für Details des Designs). Wie bereits zuvor beschrieben, erwarten wir, dass die mit den höheren Antwortkategorien konfrontierten Befragten eine

längere tägliche Fernsehdauer berichten, als diejenigen, die die Frage anhand der niedrigeren Antwortkategorien beantwortet. Unsere Ergebnisse bestätigen im Wesentlichen diese Voraussage: Insgesamt ist der Anteil der Befragten, die einen täglichen TV-Konsum von mehr als 2½ Stunden berichten, um 31 Prozentpunkte höher, wenn sie die höheren Antwortkategorien benutzen ($p < 0,001$).

Tabelle 4: Größe des Skalen-Effekts (Prozentpunkte), nach Alter und Bildungsniveau

Altersgruppe	Schulische Leistungen		Alle
	Gut	Mittel/Schlecht	
15 bis 20 Jahre	24 +	37 ***	34 ***
21 bis 60 Jahre	20 **	25 **	26 ***
61 bis 90 Jahre	21 +	41 ***	35 ***
Alle	22 ***	33 ***	31 ***

Studie 3, Frage 6-2. Angegeben ist die Prozentsatzdifferenz für den Anteil der Nennungen über 2½ Stunden Fernsehdauer in der Version mit den hohen Antwortkategorien minus des Anteils der Nennungen bei der Version mit den niedrigen Antwortkategorien.

+ < 0,10; ** $p < 0,01$; *** $p < 0,001$ für die jeweiligen Effekte; Test auf Unterschiede zwischen den Effekten: Alter nicht signifikant, Bildungsniveau $p < 0,001$; vgl. die Anmerkung zu Tabelle 1.

Studie 3 ermöglicht uns eine vergleichende Analyse von jüngeren Befragten zwischen 15 und 20 Jahren und älteren Befragten zwischen 61 und 90 Jahren und zudem einen Vergleich nach Bildungsniveau. Testet man diesen Effekt unter Berücksichtigung der besuchten bzw. absolvierten Schulart, die hier anstelle der schulischen Leistungen als Indikator verwendet wurde, um auch die Erwachsenen einbeziehen zu können, so ergibt sich ein deutlicher Effekt ($p < 0,001$): Befragte in niedrigen Bildungsgängen (oder Erwachsene, die niedrige Bildungsgänge absolviert haben) weisen einen signifikant größeren Skalen-Effekt auf (33 Prozentpunkte) als Befragte, die sich in höheren Bildungsgängen befinden bzw. diese absolviert haben (22 Prozentpunkte). Außerdem scheint der Skalen-Effekt für die jungen und die alten Befragten größer ausgeprägt zu sein, als in der Erwachsenenpopulation zwischen 21 und 60 Jahren (nicht signifikant). Scheinbar verfügen junge und alte Befragte über geringere kognitive Ressourcen, und weisen daher größere Skalen-Effekte auf; allerdings sind diese Unterschiede nicht signifikant.

3.3 Numerische Werte, die mit den Antwortkategorien assoziiert sind

Experiment 5

In einem weiteren Experiment (Studie 2) haben wir den Einfluss der numerischen Werte, die den einzelnen Antwortkategorien einer Einstellungsfrage zugeordnet sind, auf das Antwortverhalten überprüft. Wir nehmen an, dass die Befragten die den Antwortalternativen beigefügten numerischen Werte verwenden, um ein besseres Verständnis der zu Grunde liegenden evaluativen Dimension der Frage zu entwickeln. In diesem Experiment, das auf Schwarz und Mitarbeiter (1991a) zurückgeht, wurden die Befragten gebeten, den „Erfolg in ihrem Leben“ auf einer 11-stufigen Skala (vgl. Abbildung 4) zu schätzen. Die Endpunkte der Skala waren mit „überhaupt nicht erfolgreich“ und „sehr erfolgreich“ verbalisiert. Die übrigen Skalenpunkte waren nicht verbal ausgezeichnet, sondern nur durch numerische Werte markiert. Zwei verschiedene Varianten dieser Frage wurden zufällig Teilgruppen des Gesamtsamples präsentiert. Eine Hälfte der Befragten erhielt einen Fragebogen mit numerischen Werten von -5 bis +5 und die andere Hälfte der Befragten beantwortete die gleiche Frage unter Benutzung der Werte von 0 bis 10.

Basierend auf den in der Literatur berichteten Ergebnissen (Schwarz et al. 1991a) erwarteten wir eine signifikante Verschiebung der Bewertung des Erfolgs im Leben in Abhängigkeit von den verwendeten numerischen Werten. In einem selbst-administrierten Fragebogen werden die Befragten das Antwortkontinuum von -5 bis +5 derart interpretieren, dass der Wert 0 den neutralen Mittelpunkt (nicht erfolgreich) repräsentiert, wohingegen -5 das Gegenteil von Erfolg, also z.B. Misserfolg symbolisiert (Schwarz/Hippler 1995). Im Vergleich werden Befragte, die die gleiche Frage mit den Werten 0 bis 10 präsentiert bekommen, den linken extremen Punkt (also den Wert 0) als Abwesenheit von Erfolg und nicht als Misserfolg interpretieren. Daher sollte die Einschätzung des Erfolgs im Leben bei Befragten, die mit einer Skala von -5 bis +5 konfrontiert sind, in Richtung auf das positive Ende der Skala verschoben sein.

Abbildung 4: Fragebogenversionen in Experiment 5 zum Effekt der mit den Antwortalternativen assoziierten numerischen Werte, „Erfolg im Leben“

13. Was würdest du sagen: wie erfolgreich warst du in deinem Leben bisher alles in allem?

<input type="checkbox"/> -5	<input type="checkbox"/> -4	<input type="checkbox"/> -3	<input type="checkbox"/> -2	<input type="checkbox"/> -1	<input type="checkbox"/> 0	<input type="checkbox"/> +1	<input type="checkbox"/> +2	<input type="checkbox"/> +3	<input type="checkbox"/> +4	<input type="checkbox"/> +5
überhaupt nicht erfolgreich					0					sehr erfolgreich

13. Was würdest du sagen: wie erfolgreich warst du in deinem Leben bisher alles in allem?

<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	<input type="checkbox"/> 7	<input type="checkbox"/> 8	<input type="checkbox"/> 9	<input type="checkbox"/> 10
überhaupt nicht erfolgreich										sehr erfolgreich

Tabelle 5 fasst die Ergebnisse dieses Experiments zusammen. Insgesamt beobachten wir eine nennenswerte Verschiebung der Antwortverteilung in Abhängigkeit von den verwendeten numerischen Werten. Wenn Befragte mit der Skala von -5 bis +5 konfrontiert werden, ist der Anteil der Befragten, die eine Antwort oberhalb des Mittelpunktes wählen, um 13 Prozentpunkte höher ($p < 0,001$) – analog zu den in der Literatur berichteten Befunden (Schwarz et al. 1991a; Schwarz/Hippler 1995).

Auch hier benutzen wir die Größe dieses Effekts als abhängige Variable, um den Einfluss von Alter und schulischen Leistungen auf den Frage-Antwort-Prozess zu untersuchen. Weil wir vermuten, dass jüngere Kinder und Schüler mit schlechteren schulischen Leistungen geringer ausgeprägte kognitive Fähigkeiten aufweisen, rechnen wir bei ihnen mit einem größeren Effekt. Denn diese Befragten sind weniger in der Lage, der komplizierten Aufgabe beim Aufbau einer kognitiven Repräsentation der evaluativen Dimension dieser Frage allein anhand der verbalisierten Endpunkte des Antwortkontinuums nachzukommen. Daher werden sie sich bei der Entwicklung des Frageverständnisses und der Generierung einer Antwort stärker von den numerischen Werten beeinflussen lassen.

Tabelle 5: Effekte der den Antwortalternativen beigefügten numerischen Werte (Prozentpunkte), nach Alter und schulischen Leistungen

Altersgruppe	Schulische Leistungen		Alle
	Gut	Mittel/Schlecht	
13 bis 15 Jahre	9 *	18 ***	16 ***
16 bis 17 Jahre	8 *	13 ***	12 ***
18 bis 21 Jahre	8 **	12 ***	11 ***
Alle	9 ***	15 ***	13 ***

Studie 2, Frage 13. Angegeben ist die Prozentsatzdifferenz: Anteil der Befragten über dem Mittelpunkt der Skala in der Version -5/+5 minus des gleichen Anteils in der Version 0/10.

* < 0,05; *** $p < 0,001$ für die jeweiligen Effekte; Test auf Unterschiede zwischen den Effekten: Alter $p < 0,01$, schulische Leistungen $p < 0,001$; vgl. die Anmerkung zu Tabelle 1.

Wie erwartet weisen ältere Befragte einen geringeren Antworteffekt auf (11 Prozentpunkte) als jüngere Befragte (16 Prozentpunkte; $p < 0,01$). Weiterhin beobachten wir einen Unterschied zwischen Befragten mit mittleren und schlechten schulischen Leistungen (15 Prozentpunkte) und Schülern mit besserer schulischer Leistungsfähigkeit (9 Prozentpunkte; $p < 0,001$). Beide Befunde gehen mit den theoretischen Erwartungen konform. Eine detaillierte Inspektion der Tabelle ergibt weiter, dass wir vor allem für die Befragten mit durchschnittlicher und schlechterer schulischer Leistungsfähigkeit einen

ausgeprägten Alterseffekt nachweisen können. Hingegen finden wir keinen Alterseffekt in der Gruppe mit überdurchschnittlichen schulischen Leistungen.

3.4 Fragereihenfolgen-Effekte

Bisher haben geringere kognitive Fähigkeiten zu stärkeren Antworteffekten geführt (Antwortreihenfolgen-Effekte, Skalen-Effekte und Effekte der numerischen Werte an den Antwortalternativen). Im Vergleich dazu wirkt die kognitive Kapazität mit umgekehrtem Vorzeichen auf die Höhe der Fragereihenfolgen-Effekte. Da diese auf der Berücksichtigung von zuvor gestellten Fragen bei der Interpretation und Beantwortung der aktuellen Frage beruhen, müssen Befragte, die zuvor wahrgenommenen Informationen speichern und als relevant für die Beantwortung der aktuellen Frage wahrnehmen. Diese kognitive Aufgabe können jüngere Befragte und Befragte mit schlechteren schulischen Leistungen weniger gut bewältigen, weshalb der Inhalt der zuvor gestellten Frage bei ihnen eine geringere Rolle für das Frageverständnis der aktuellen Frage und deren Beantwortung spielt. Entsprechend sind die Fragereihenfolgen-Effekte bei den Befragten mit geringeren kognitiven Kapazitäten geringer ausgeprägt.

Experiment 6

Im ersten Experiment zur Fragereihenfolge (Studie 2) wurden die Befragten gebeten, ihre Chancen auf einen Arbeitsplatz nach der Schule im Vergleich zu anderen Jugendlichen in Ostdeutschland und Westdeutschland zu schätzen. Dabei wurden jedem Befragten (es handelt sich nur um Westdeutsche im Bundesland Bayern) zwei Fragen gestellt: In der ersten sollten die Chancen im Vergleich zu anderen Jugendlichen in Westdeutschland geschätzt werden; in der zweiten Frage sollte ein entsprechender Vergleich mit Jugendlichen in Ostdeutschland vorgenommen werden (Abbildung 5). Diese beiden Fragen wurden in der Originalreihenfolge (der Vergleich mit Jugendlichen in Westdeutschland zuerst) und in umgekehrter Reihenfolge (der Vergleich mit ostdeutschen Jugendlichen zuerst) gestellt.

Abbildung 5: Fragebogenversionen in Experiment 6 zur Fragereihenfolge, „Chancen auf einen Arbeitsplatz nach der Schule“, Originalversion (rechts) und gedrehte Version (links)

11. Und wenn du dich mit den Jugendlichen in OST-Deutschland allgemein vergleichst, hast du dann deiner Meinung nach bessere Chancen auf einen Arbeitsplatz, schlechtere Chancen oder etwa die gleichen Chancen auf einen Arbeitsplatz?

- bessere Chancen
- etwa die gleichen Chancen
- schlechtere Chancen

12. Und wenn du dich mit den Jugendlichen in WEST-Deutschland allgemein vergleichst, hast du dann deiner Meinung nach bessere Chancen auf einen Arbeitsplatz, schlechtere Chancen oder etwa die gleichen Chancen auf einen Arbeitsplatz?

- bessere Chancen
- etwa die gleichen Chancen
- schlechtere Chancen

11. Und wenn du dich mit den Jugendlichen in WEST-Deutschland allgemein vergleichst, hast du dann deiner Meinung nach bessere Chancen auf einen Arbeitsplatz, schlechtere Chancen oder etwa die gleichen Chancen auf einen Arbeitsplatz?

- bessere Chancen
- etwa die gleichen Chancen
- schlechtere Chancen

12. Und wenn du dich mit den Jugendlichen in OST-Deutschland allgemein vergleichst, hast du dann deiner Meinung nach bessere Chancen auf einen Arbeitsplatz, schlechtere Chancen oder etwa die gleichen Chancen auf einen Arbeitsplatz?

- bessere Chancen
- etwa die gleichen Chancen
- schlechtere Chancen

Wir erwarten, dass die Fragereihenfolge einen moderierenden Effekt auf die perzipierte Bedeutung der Fragen ausübt: Vergleichen Befragte sich selbst zunächst mit Jugendlichen in Ostdeutschland, sollten beim anschließenden Vergleich mit Jugendlichen in Westdeutschland bessere Einschätzungen gemessen werden. Wir vermuteten, dass die ggf. wahrgenommene Unterlegenheit der eigenen Person im Vergleich zu den westdeutschen Altersgenossen vor dem Hintergrund der größeren Distanz zu den Ostdeutschen kleiner erscheint; die eigene Lage wird dadurch positiver berichtet.

Tabelle 6: Größe des Fragenreihenfolgen-Effekts, „Chance auf einen Arbeitsplatz nach der Schule“ (Prozentpunkte), nach Alter und schulischen Leistungen

Altersgruppe	Schulische Leistungen		Alle
	Gut	Mittel/Schlecht	
13 bis 15 Jahre	2	0	0
16 bis 17 Jahre	7	-1	1
18 bis 21 Jahre	13	6	10
Alle	7 **	5 **	6 **

Studie 2, Frage 11. Angegeben ist die Prozentsatzdifferenz des Anteils der Befragten, die eine gute Chance auf einen Arbeitsplatz erwarten in der gedrehten Version minus des entsprechenden Anteils in der Originalversion.

* $p < 0,05$; ** $p < 0,01$ für den jeweiligen Effekt; Test auf Unterschiede zwischen den Effekten: Alter $p < 0,01$, schulische Leistungen $p < 0,11$; vgl. die Anmerkung zu Tabelle 1.

Insgesamt finden wir einen moderaten erwartungskonformen Fragereihenfolgen-Effekt von 6 Prozentpunkten ($p < 0,01$): Wenn die befragten Kinder und Jugendlichen sich zunächst mit den ostdeutschen Altersgenossen vergleichen und dann erst mit den Westdeutschen, dann ist der Anteil derjenigen, die sich selbst im Vergleich mit den Westdeutschen bessere Chancen auf einen Arbeitsplatz zumessen, um 6 Prozentpunkte höher, als wenn der Vergleich mit den Ostdeutschen nach dem Vergleich mit den Westdeutschen erfragt wird. Auch hier benutzen wir die Größe dieses Effektes als abhängige Variable und setzen sie mit der kognitiven Kapazität in Beziehung: Bei älteren Befragten zeigen sich signifikante Unterschiede der Beantwortung der Frage nach den Chancen auf einen Arbeitsplatz im Vergleich zu westdeutschen Jugendlichen, wohingegen bei jüngeren Befragten kein signifikanter Reihenfolgen-Effekt ausweisbar ist ($p < 0,01$). Der Effekt der schulischen Leistungen deutet zwar in die vorausgesagte Richtung, erreicht aber nicht das statistische Signifikanzniveau ($p < 0,11$).

Experiment 7

In einem zweiten Fragereihenfolgen-Experiment (Studie 2) wurden die Befragten gebeten, ihre Zufriedenheit mit ihrem Leben im Allgemeinen und ihre Zufriedenheit mit ihrer Situation in der Klasse anzugeben. In diesem Experiment haben wir die Logik eines klassischen Designs von Schwarz und Mitarbeiter (1991b; vgl. auch Sudman et al. 1996) auf jugendliche Befragte zugeschnitten. Je die Hälfte der Befragten erhielt eine von zwei experimentellen Fragebogenversionen zugewiesen: Version 1 enthielt die beiden Fragen in

der Originalreihenfolge, in der zunächst nach der Zufriedenheit mit dem Leben allgemein und dann erst nach der Zufriedenheit mit der Klasse gefragt wurde. Version 2 stellte die beiden Fragen in umgekehrter Reihenfolge (Abbildung 6). Wir erwarten, dass die Beurteilung der Situation in der Klasse negativer ausfällt, wenn die Frage an zweiter Stelle positioniert ist. Wenn die Kinder und Jugendlichen zunächst nach der Zufriedenheit mit dem Leben insgesamt gefragt werden und dann erst nach der Zufriedenheit mit ihrer Klasse, dann wird die Klasse im Kontrast zum Leben allgemein stärker als der Ort des Lernens und der Prüfungen aufgefasst, und weniger als Ort, wo man Freunde treffen oder gesellig sein kann. Wird hingegen nach der Klasse an erster Stelle gefragt, interpretieren die Befragten die Klasse als Zweiklang aus formalisierten Lernprozessen und sozialen Aktivitäten. Dies führt zu einer besseren Bewertung der Klasse, wenn die betreffende Frage an erster Stelle steht.

Abbildung 6: Fragebogenversionen in Experiment 7 zur Fragereihenfolge, „Zufriedenheit mit der Klasse“, Originalversion (links) und gedrehte Version (rechts)

1. Wie zufrieden bist du mit deinem Leben im Allgemeinen?

- Ich bin sehr zufrieden
- Ich bin eher zufrieden
- Ich bin eher unzufrieden
- Ich bin sehr unzufrieden

2. Wie zufrieden bist du mit deiner Klasse?

- Ich bin sehr zufrieden
- Ich bin eher zufrieden
- Ich bin eher unzufrieden
- Ich bin sehr unzufrieden

1. Wie zufrieden bist du mit deiner Klasse?

- Ich bin sehr zufrieden
- Ich bin eher zufrieden
- Ich bin eher unzufrieden
- Ich bin sehr unzufrieden

2. Wie zufrieden bist du mit deinem Leben im Allgemeinen?

- Ich bin sehr zufrieden
- Ich bin eher zufrieden
- Ich bin eher unzufrieden
- Ich bin sehr unzufrieden

Insgesamt finden wir in den Daten einen moderaten Fragereihenfolgen-Effekt von 7 Prozentpunkten ($p < 0,001$). Befragte, die zunächst nach der Klasse und dann erst nach ihrem Leben insgesamt befragt wurden, beurteilen die Situation in der Klasse entsprechend besser, als wenn sie die Fragen in umgekehrter Reihenfolge präsentiert bekommen.

Tabelle 7: Größe des Fragereihenfolgen-Effekts, „Zufriedenheit mit der Klasse“ (Prozentpunkte), nach Alter und schulischer Leistung

Altersgruppe	Schulische Leistungen		Alle
	Gut	Mittel/Schlecht	
13 bis 15 Jahre	9 *	4 *	5 *
16 bis 17 Jahre	6 +	8 ***	7 ***
18 bis 21 Jahre	11 ***	7 *	9 ***
Alle	8 ***	7 ***	7 ***

Studie 2, Frage 2 (ohne Berufsschüler). Angegeben ist die Prozentsatzdifferenz zwischen dem Anteil der Befragten, die sich in der gedrehten Version „sehr zufrieden“ oder „eher zufrieden“ mit der Klasse zeigen minus des entsprechenden Anteils in der Originalversion.

+ < 0,10; * $p < 0,05$; *** $p < 0,001$ für den jeweiligen Effekt; Test auf Unterschiede zwischen den Effekten: Alter $p < 0,001$, schulische Leistungen nicht signifikant; vgl. die Anmerkung zu Tabelle 1.

Auch diesen Fragereihenfolgen-Effekt haben wir nach Altersgruppen und schulischer Leistung differenziert und kommen zu folgenden Ergebnissen: Jüngere Befragte weisen im Vergleich zu älteren Jugendlichen einen signifikant kleineren Fragereihenfolgen-Effekt auf ($p < 0,001$). Außerdem können wir für Schüler mit besseren schulischen Leistungen leicht größere Fragenreihenfolgen-Effekte berechnen – allerdings sind diese Differenzen nicht statistisch signifikant.

Nimmt man die Ergebnisse der beiden Experimente zur Fragereihenfolge zusammen, so unterstützen die Befunde die Hypothese, wonach sich mit steigendem Alter der Umfang des relevanten Kontextes, der bei der Beantwortung einer Frage berücksichtigt wird, vergrößert. Nicht bestätigt werden kann hingegen die These, wonach die intellektuellen Fähigkeiten innerhalb einer Altersgruppe (gemessen durch die schulischen Leistungen) auf den Umfang des wahrgenommenen Kontextes einer Frage einwirken.

4. Diskussion

Betrachtet man die Ergebnisse aller Experimente zusammen, so zeigt sich ein durchgängiger Alterseffekt auf die Antworteffekte in der erwarteten Richtung, der mit Ausnahme von Experiment 4 auch statistisch gesichert ist. Sowohl bei der Antwortreihenfolge, bei den Skalen-Effekten und bei den Effekten der numerischen Werte sowie bei der Fragereihenfolge finden sich die erwarteten alterabhängigen Unterschiede in der Effektgröße. Mit Einschränkungen gilt dies auch für die schulischen Leistungen: Hier weisen die beobachteten Effekte – mit Ausnahme von Experiment 1 – in die durch die Hypothesen vorausgesagte Richtung. Außerdem ergibt die statistische Prüfung bei den

sagte Richtung. Außerdem ergibt die statistische Prüfung bei den Skalen-Effekten, bei den Effekten der numerischen Werte an den Antwortalternativen und bei der Antwortreihenfolge – bei einem von zwei Experimenten – gesicherte Unterschiede zwischen den Leistungsgruppen. Lediglich bei der Fragereihenfolge ergeben sich keine statistisch signifikanten Befunde.

Tabelle 8: Ergebnisse im Überblick, „Unterscheiden sich die Effekte in der erwarteten Richtung?“, Signifikanz der Effektunterschiede

	Indikatoren für die kognitive Kapazität	
	Alter	Schulische Leistungen
Effekt der Antwortreihenfolge		
Experiment 1	Ja +	Nein
Experiment 2	Ja **	Ja *
Skalen-Effekt		
Experiment 3	Ja **	Ja *
Experiment 4	Ja	Ja ***
Effekt der numerischen Werte		
Experiment 5	Ja **	Ja ***
Effekt der Fragereihenfolge		
Experiment 6	Ja **	Ja
Experiment 7	Ja **	Ja

+ p < 0,10; * p < 0,05; ** p < 0,01; *** p < 0,001.

Inwieweit die wenigen nicht erwartungskonformen Befunde auf methodische Schwächen unseres Designs oder auf falsche Hypothesen zurückzuführen sind, kann an dieser Stelle nicht entschieden werden. Die im Großen und Ganzen die Hypothese bestätigenden Ergebnisse ermutigen uns aber, in weiteren experimentellen Studien den Einfluss der kognitiven Fähigkeiten von Kindern und Jugendlichen auf die Beantwortung von Fragebögen zu untersuchen.⁶ Die vorliegenden Ergebnisse bilden zwar noch kein umfassendes Bild der kognitiven Prozesse dieser Befragtengruppen bei der Beantwortung von standardisierten Fragebögen, aber es ergibt sich eine Reihe von Hinweisen:

⁶ Derzeit ist eine weitere Untersuchung mit 4.500 befragten Kindern und Jugendlichen im Alter zwischen 10 und 21 Jahren im Feld, in die wir wiederum Experimente zu Antwort-Effekten integriert haben. Zudem haben wir versucht, die unabhängige Variable „schulische Leistungen“ durch Schulnoten in drei Hauptfächern zu erheben. Ergebnisse dieser Experimente liegen noch nicht vor.

Kinder und jüngere Jugendliche im Alter von 10 bis etwa 14 Jahre entwickeln aufgrund ihrer noch limitierten kognitiven Fähigkeiten ein eingeschränktes Frageverständnis; gleichzeitig lassen sie sich bei der Beantwortung der Frage stärker durch die vorgegebenen Antwortalternativen leiten. Es kompensieren sich hier also zwei gegenläufige Prozesse. Überspitzt könnte man sagen: Die jüngeren Befragten beantworten auch komplizierte Fragen, weil sie ihre noch limitierten kognitiven Fähigkeiten davor bewahrten zu erkennen, dass diese Fragen für sie ein Problem darstellen sollten. Dies mag im Übrigen auch eine Erklärung sein, warum der Item-Nonresponse bei Kindern zwar stärker ausgeprägt ist als bei Jugendlichen, aber auch kein übermäßig großes Problem darstellt (Borgers/Hox 2001; Fuchs 2003). Ältere Befragte betten Einzelfragen hingegen stärker in den Fragefluss bzw. -kontext ein, sie entwickeln ein breiteres Bild des Frageinhalts und sind daher bei der Generierung einer Antwort in einer anderen (besseren) Situation: Sie können die Antwort ohne (oder mit geringerem) Rückgriff auf die Spannweite des vorgegebenen Antwortkontinuums und auf die den Antwortvorgaben beigefügten numerischen Werte produzieren, und sie können allen Antwortalternativen die gleiche Aufmerksamkeit widmen. Damit scheint sich zumindest ansatzweise unsere generelle Annahme zu bestätigen, dass jüngere Befragte im Verhältnis zum Fragebogen in einer schwächeren, weniger autonomen Position verbleiben. Überspitzt könnte man sagen, dass jüngere Befragte einen möglichst stromlinienförmigen Weg durch die Befragung suchen, ohne eigene Spuren (= Anhaltspunkte für den „wahren Wert“) zu hinterlassen. Ob es sich dabei um ein Verhalten handelt, das Krosnick (1991, 1992) als „Satisficing“ bezeichnet hat, oder ob die befragten Kinder einfach nur versuchen, bei der anspruchsvollen Aufgabe der Beantwortung eines Fragebogens möglichst gut zu kooperieren (also versuchen, den Fragebogen auch bei Unklarheiten oder Problemen möglichst vollständig auszufüllen, vgl. Holaday/Turner-Henson 1989), kann an dieser Stelle noch nicht entschieden werden.

Einige wichtige Einschränkungen sind bei der Interpretation der Ergebnisse zu berücksichtigen: In unseren Analysen wurden Alter und schulische Leistungen als Indikatoren für die kognitive Kapazität verwendet. Wir sind uns jedoch bewusst, dass der Prozess der Entwicklung der kognitiven Kapazitäten nicht einer linearen altersbedingten Logik folgt. Vielmehr mögen manche Kinder bereits in jüngeren Jahren ein kognitives Stadium vorstoßen, dass andere erst zu einem späteren Zeitpunkt in ihrem Leben erreichen (Dockrell et al. 2000). Außerdem stehen Bildung und schulische Leistung nicht nur für kognitive Kapazität, sondern könnten ebenso Motivation und Leistungsbereitschaft repräsentieren. Hier kann erst eine Weiterentwicklung des Konzeptes der kognitiven Kapazität und seiner Operationalisierung zur Klärung beitragen (vgl. die Arbeiten zu den kognitiven Kapazitäten von alten und hochbetagten Befragten: Knäuper et al 1997; Knäuper 1999;

Schwarz/Knäuper 1999). Zu denken ist hierbei etwa an separate Messungen von Arbeitsgedächtnis und Aufmerksamkeit (vgl. Schwarz et al. 1998).

Weiter haben wir mit den referierten Ergebnissen nur einzelne Ausschnitte des Frage-Antwort-Prozesses untersuchen können. Insbesondere die Phase des Formatierens und des Editierens ist bisher weitgehend unberücksichtigt geblieben. Zudem stellt sich die Frage, ob die hier berichteten Unterschiede zwischen Kindern, Jugendlichen und volljährigen Befragten auch bei anderen Themen und mit anderen Befragungsmodes (mündlich-persönliche oder telefonische Befragung) repliziert werden können. Eine Generalisierung auf Befragungen von Kindern aller Art ist daher nicht zulässig.

Aus der jeweiligen Stichprobe der zugrunde liegenden drei Studien ergab sich für unsere Experimente eine Altersbegrenzung auf Kinder ab 10 Jahre. Eine Übertragung unserer Befunde auf Kinder jüngeren Alters ist daher nicht gerechtfertigt. Das Interesse der Umfrageforschung an Grundschulkindern wird aber vermutlich steigen (vgl. etwa das DJI-Kinderpanel), daher wäre eine Ausdehnung unserer Forschungslinie auf jüngere Kinder sinnvoll.

Weiter steht außer Frage, dass sich mit zunehmendem Alter auch die sozialen Kompetenzen der Kinder – und damit ihr Verständnis für die spezifische Situation „Befragung“ – vergrößert. Kohlberg (1976) hat die Bedeutung der sozialen Kompetenzen betont und in zukünftigen Studien sollten die kognitiven und sozialen Kompetenzen separat voneinander gemessen und als unabhängige Variablen benutzt werden.

In einem weiteren Schritt wäre es sinnvoll, die kognitiven Prozesse von Kindern und Jugendlichen mit denen von Erwachsenen an einem einheitlichen Datensatz zu vergleichen. Die Ergebnisse von Experiment 4 legen die Vermutung nahe, dass jüngere Befragte allgemein stärker von Antworteffekten betroffen sind als ein Querschnitt der Erwachsenenpopulation. Für eine detaillierte Analyse dieses Unterschieds fehlen uns jedoch bisher die Daten und wir sollten daher zukünftige Experimente so anlegen, dass Kinder, Jugendliche und Erwachsene miteinander vergleichbar sind und wir anhand der Größe der Antworteffekte entscheiden können, ab welchem Alter sich Kinder und Jugendliche bei der Beantwortung eines Fragebogens wie Erwachsene verhalten.

Weiter sollten Merkmale der einzelnen Fragen berücksichtigt werden, um aufzuzeigen, bei welchen Fragetypen die begrenzten kognitiven Kapazitäten der Kinder und jüngeren Jugendlichen zu Schwierigkeiten führen: Unsere Befunde zum Einfluss von Eigenschaften der jeweiligen Frage auf den Item-Nonresponse bei Kindern und Jugendlichen unterschiedlichen Alters und unterschiedlichen schulischen Leistungsniveaus (Fuchs 2003) verdeutlichen, dass die Probleme des Item-Nonresponse vor allem dann auftreten, wenn

die Fragen aufgrund der formalen Eigenschaften (z.B. Mehrfachantwort, Item-Batterie) oder aufgrund von inhaltlichen Anforderungen als schwierig einzustufen sind. In zukünftigen Experimenten sollten daher die hier geschilderten Experimente für Fragen auf unterschiedlichem Schwierigkeitsniveau repliziert werden. Auch Verständnisschwierigkeiten mit den Frageformulierungen sollten untersucht werden, denn möglicher Weise sind die beobachteten Effekte auch auf nicht altersgemäße Frageformulierungen zurückzuführen.

Trotz dieser Einschränkungen und der weiteren Forschungsfragen geben unsere Befunde Hinweise darauf, dass sich Kinder und Jugendliche hinsichtlich der Abarbeitung des Frage-Antwort-Prozesses von erwachsenen Befragten unterscheiden. Um daraus Konsequenzen für die Fragebogenkonstruktion zu ziehen, ist es derzeit wohl noch zu früh, weil wir nur über ein begrenztes Verständnis der Besonderheiten des Frage-Antwort-Prozesses bei Kinder und Jugendlichen verfügen. Aber zumindest scheint gesichert, dass wir den für erwachsene Befragte erarbeiteten methodischen Kenntnisstand nicht ohne weitere Prüfung auf Kinder und Jugendliche übertragen können. Das klingt wie eine methodische Binsenweisheit, aber zumindest erhalten wir auf Basis der referierten Ergebnisse eine Ahnung von Art und Umfang der Unterschiede bei der Verarbeitung von Fragebogenfragen.

Literatur

Amato, P./Ochiltree, G., 1987: Interviewing children about their families: a note on data quality. In: *Journal of Marriage and the Family*, 49, 669-675.

Benson, J./Hocevar, D., 1985: The impact of item phrasing on the validity of attitude scales for elementary school children. In: *Journal of Educational Measurement*, 22, 231-240.

Borgers, N./Hox, J., 2000: Reliability of responses in self-administered questionnaires: Research with children. In: Blasius, J./Hox, J./de Leeuw, E./Schmidt, P. (Hrsg.), *Social Science Methodology in the New Millennium. Proceedings of the Fifth International Conference on Logic and Methodology*, October 3-6, Cologne, Germany.

Borgers, N./Hox, J., 2001: Item nonresponse in questionnaire research with children. In: *Journal of official statistics*, 17, 321-335.

Borgers, N./de Leeuw, E./Hox, J., 2000: Children as respondents in survey research: Cognitive development and response quality. In: *Bulletin de Methodologie Sociologique (BMS)*, 66, 2000, 60-75.

Dockrell, J./Lewis, A./Lindsay, G., 2000: Researching childrens' perspectives: a psychological dimension. In: Lewis, A./Lindsay, G. (Hrsg.), *Researching childrens' perspectives*. Buckingham, 46-58.

Fuchs, M., 2003: Data quality in surveys among the youth. Item non-response and response errors. Vortrag anlässlich der internationalen Fachtagung "Methoden und Techniken der Kindheitsforschung", Schloss Hohenkammer, April 2003.

Fuchs, M./Lamnek, S./Luedtke, J., 1999: *Tatort Schule. Gewalt an Schulen 1994 – 1999*. Opladen.

Fuchs, M./Lamnek, S./Wiederer, R., 2001: *Querschläger. Jugendliche zwischen rechter Ideologie und Gewalt*. Opladen.

Groves, R., 1989: *Survey errors and survey costs*. New York.

Hershey, M./Hill, D., 1976: Positional response set in pre-adult socialisation surveys. In: *Social Science quarterly*, 56, 707-714.

Hippler, H.-J./Schwarz, N., 1992: The impact of administration mode on response effects in surveys. *ZUMA-Arbeitsberichte* 92/14. Mannheim, Germany.

Holaday, B./Turner-Henson, A., 1989: Response effects in surveys with school-age children. In: *Nursing Research*, 38, 248-250.

Knäuper, B., 1999: The impact of age and education on response order effects in attitude measurement. In: *Public Opinion Quarterly*, 63, 347-370.

Knäuper, B./Belli, R. Hill, D./Herzog, R., 1997: Question difficulty and respondent's cognitive ability: the effect on data quality. In: *Journal of Official Statistics*, 13, 181-199.

Kohlberg, L., 1976: Moral stages and moralization. The cognitive-developmental approach. In: Lickona, T., (Hrsg.), *Moral development and behavior. Theory, research and social issues*. New York, 31-53.

Krosnick, J., 1991: Response strategies for coping with the cognitive demands of attitude measures in surveys. In: *Applied Cognitive Psychology*, 5, 213-236.

Krosnick, J., 1992: The impact of cognitive sophistication and attitude importance on response order and question order effects. In: Schwarz, N./Sudman, S. (Hrsg.), *Context effects in social and psychological research*. New York, pp. 203-218.

Krosnick, J./Alwin, D., 1987: An evaluation of cognitive theory of response-order effect in survey measurement. In: *Public Opinion Quarterly*, 51, 201-219.

De Leeuw, E./Otter, M., 1995: The reliability of children's responses to questionnaire items: question effects in children's questionnaire data. In: Hox, J./van der Meulen, B./Kanssens, J./ter Laak, J./Tavecchio, L. (Hrsg.), *Advances in family research*. Amsterdam, 251-257.

Lyberg, L./Biemer, P./Collins, M./de Leeuw, E./Dippo, C./Schwarz, N./Trewin/D. (Hrsg.), 1997: *Survey measurement and process quality*. New York.

Marsh, H., 1986: Negative item bias in rating scales for preadolescent children: a cognitive-developmental phenomenon. In: *Developmental Psychology*, 22, 37-49.

Miller, W., 1976: Preschoolers' attitude toward libraries: can they be measured? In: *The Library Quarterly*, März 1976, 151-153.

Naplava, T./Oberwittler, D., 2002: Methodeneffekte bei der Messung selbstberichteter Delinquenz von männlichen Jugendlichen. In: *Monatsschrift für Kriminologie und Strafrechtsreform*, 85, 401-423.

Piaget, J., 1948: *The moral judgment of the child*. Glencoe, Ill.

Schuman, H./Presser, S., 1981: Questions and answers in attitudes surveys. Experiments on question form, wording, and context. Thousand Oaks.

Schwarz, N./Hippler, H.-J./Deutsch, B./Strack, F., 1985: Response scales. Effects of category range on reported behaviour and comparative judgements. In: *Public Opinion Quarterly*, 49, 388-395.

Schwarz, Norbert/Hippler, H.-J./Noelle-Neumann, E., 1992: A cognitive model of response order effects in survey measurement. In: Schwarz, N./Sudman, S. (Hrsg.), *Context effects in social and psychological research*. New York, 187-201.

Schwarz, N./Hipper, H.-J., 1995: The numeric values of rating scales: a comparison of their impact in mail surveys and telephone interviews. In: *International Journal of Public Opinion Research*, 7, 72-74.

Schwarz, N./Knäuper, B., 1999: Cognition, aging, and self-reports. In: Park, D./Schwarz, N. (Hrsg.), *Cognitive aging. A student primer*. Philadelphia, 233-252.

Schwarz, N./Knäuper, B./Hippler, H.-J./Noelle-Neumann, E./Clark/L., 1991a: Rating scales. Numeric values may change the meaning of scale labels. In: *Public Opinion Quarterly*, 55, 570-582.

Schwarz, N./Knäuper, B., Park, D., 1998: *Cognition, aging, and self-reports*. Philadelphia, PA.

Schwarz, N./Strack, F./Mai, H., 1991b: Assimilation and contrast effects in part-whole question sequences. A conversational logic analysis. In: *Public Opinion Quarterly*, 55, 3-23.

Scott, J., 1997: Children as respondents: Methods for improving data quality. In: Lyberg, L./Biemer, P./Coolins, M./de Leeuw, E./Dippo, C./Schwarz, N./Trewin, D. (Hrsg.), *Survey measurement and process quality*. Wiles, New York, 331-350.

Sudman, S./Bradburn, N./Schwarz, N., 1996: *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco.

Tourangeau, R./Rips, L./Rasinski, K., 2000: *The psychology of survey response*. New York.

Villancourt, P., 1973: Stability of children's survey responses. In: *Public Opinion Quarterly*, 37, 373-387.

Wänke, M./Schwarz, N., 1997: Reducing question order effects: The operation of buffer items. In: Lyberg, L./Biemer, P./Collins, M./de Leeuw, E./Dippo, C./Schwarz, N./Trewin, D., (Hrsg.), *Survey measurement and process quality*. New York, 115-140.

Korrespondenzadresse

PD Dr. Marek Fuchs

Universität Kassel

Fachbereich 5

D-34109 Kassel

email: marek.fuchs@uni-kassel.de