Journal of Official Statistics, Vol. 21, No. 4, 2005, pp. 701–725

Children and Adolescents as Respondents. Experiments on Question Order, Response Order, Scale Effects and the Effect of Numeric Values Associated with Response Options

Marek Fuchs[1]

The results presented in this article are relevant to the general question to what extent the available methodological knowledge regarding the question–answer process applies also to children and adolescents. Based on the assumption that children lack full cognitive functioning due to their cognitive-developmental stage, we predicted a less sophisticated question–answer process for these respondents. In order to assess this hypothesis we have focused on age and educational achievement as explanatory factors, while the size of response effects was treated as an indicator for the cognitive processes in the respondents' minds while preparing to answer a survey question.

In sum, the size of response order effects as well as that of scale effects and that of the effects of numeric values associated with response categories decreases with age. In general, in each age-group comparison the response effects reach the highest level among the youngest group – that is assumed to be cognitively less developed. At the same time, younger respondents are more likely to ignore contextual information when decoding the question than older respondents. With some restrictions this holds also for the differences between the size of the response effects for respondents with good school achievement on the one hand and respondents with intermediate or poor school achievement on the other.

Key words: Measurement error; response effect; field experiment.

1. Introduction and Hypothesis

Both theory of the question–answer process and the rules for designing a "good" questionnaire are well-developed. As a result, experts as well as novices in survey research can easily access many comprehensive books on data quality (e.g., Groves 1989; Lyberg et al. 1997) and questionnaire design (e.g., Schuman and Presser 1981; Sudman et al. 1996; Tourangeau et al. 2000). However, the methodological knowledge described in the literature is predominantly developed for surveys in the adult population.

By contrast, in recent years some attention has been given to the special needs of surveys among the elderly (Herzog and Rodgers 1992; Knäuper et al. 1992) and other special populations like men, women, the ill, and elites (Gubrium and Holstein 2002). These efforts at tailoring the well-developed survey technique to the specifics of particular populations are motivated by two needs. First, when conducting surveys in the general population researchers face the question whether or not differences between the data of certain subpopulations are due to differences in their specific attitudes and behaviors.

[1] University of Kassel, Department 5, D-34109 Kassel, Germany. Email: marek.fuchs@uni-kassel.de
Acknowledgments: The author is grateful to four anonymous reviewers and to the editor of the journal for many helpful comments on a previous version of this article.

This question arises on the basis of the hypothesis that differences in the substantial results might also occur because of varying degrees of measurement error among the age groups. Second, researchers need to collect data of high quality in these special populations, and thus they need to address the specific presuppositions of survey measurement in these populations in order to tailor the well-developed survey methodology accordingly.

Recently, many surveys have addressed samples of children and adolescents. Today, their behavior, attitudes and beliefs are predominantly collected directly from them. By contrast, proxy-reporting from parents or other caretakers is no longer seen as a sufficient technique of data collection (Borgers et al. 2000). Children and adolescents are assumed to be competent respondents who can speak for themselves and who are often more knowledgeable about their own activities. Collecting data directly from them will likely improve data quality (Hess et al. 1998). However, the literature has shown that children and adolescents respond better to standardized questionnaires based on limited cognitive capacities. Assuming a negative effect of the limited cognitive skills on the question–answer process (Scott 1997), the quality of data obtained from children and adolescents has been questioned.

In the literature on the psychology of the survey interview, the question–answer process is seen as a complex interaction of the respondent with – among other elements of a survey – several characteristics of a question. In this perspective, the respondent completes several tasks when generating a response to a particular item: understanding the meaning of the question, retrieving information that is relevant in the process of finding an answer, generating a response, formatting the response according to the response options provided, and editing the response (see, e.g., Sudman et al. 1996 for details). In this process, the respondent considers the question wording and the response categories in addition to surrounding information such as previous questions (see Tourangeau et al. 2000 for a comprehensive summary). Assuming that their cognitive skills are still developing, "one may wonder whether children can adequately perform these difficult tasks" (DeLeeuw and Otter 1995, p. 251).

For the purpose of conceptualizing human cognitive development, several theoretical approaches are available (e.g., Goswami 2002). On the basis of theories of cognitive and moral growth (Piaget 1948; Kohlberg 1976), survey researchers assume that children aged ten and older are capable of answering survey questions (see Scott 1997 and Borgers et al. 2000 for reviews; Benson and Hocevar 1985 address elementary school children). It is assumed that children of this age have reached the stage of formal operations (Piaget 1948) or – in Kohlberg's (1976) terminology – the first stage in the conventional level of moral development. Thus they are able and willing to follow rules independently of others being present and to think about issues in a more abstract fashion. Demetriou and colleagues (2002) extended the classic developmental model by incorporating the experimental line of research as well as the psychometric research on human development. According to their model the human mind is organized in three layers, which must be distinguished from developmental stages: The basic layer involves general processing potentials; the second layer involves specialized capacity spheres, which in turn involve a system of cognitive functioning and abilities specializing in the representation and processing of the different aspects of their environment; the third layer involves systems underlying self-monitoring, self-representation, and self-regulation. On the basis of an extensive review of

the literature, Demetriou and colleagues (2002) come to the conclusion that, with human development from childhood to early adulthood, general processing potentials become faster and more efficient, and the specialized capacity spheres become increasingly able to deal with complex and counterintuitive problems. Also, both self-awareness and self-regulation become stronger (Demetriou 2000; Case 1992; Demetriou et al. 1993 and other sources cited by Demetriou et al. 2002).

Even though it is assumed that development does not follow a strict age-related logic for all individuals, it is obvious that 10-year-old children and 17-year-old adolescents differ to a large extent with respect to the cognitive skills mentioned above. With respect to the aforementioned question–answer process, it is questionable whether young children are capable of keeping in mind all the information provided by a survey item simultaneously while processing each single question and the questionnaire as a whole. It is also unknown to what extent they process and link the information provided differently from older respondents.

According to the theoretical reasoning discussed above, human development cannot be seen as one discrete step from childhood to adulthood. One should rather view the cognitive development of children and adolescents as a process that covers an extensive period of time. Thus, children and adolescents cannot be addressed as one homogeneous group. Instead, one should differentiate between children and adolescents of different ages and developmental stages. Because cognitive skills correlate with the ability to respond to a survey appropriately, one can speculate that younger respondents apply a less sophisticated question–answer process when responding to a survey question. On the basis of this reasoning, the question arises to what extent children, adolescents and adults differ in terms of the size of response effects and other data quality indicators that can be used as outcome measures for the quality of the question–answer process.

However, little empirical work has been done on the differences between adults and underaged respondents of different age groups (see Miller 1976 for a funny parody of methodological research on surveys of children). So far, mostly secondary analysis of existing data sets has been conducted in order to assess data quality indicators in surveys with children and adolescents, e.g., item nonresponse and response stability. On the basis of a panel study on political socialization conducted in three waves, Villancourt (1973) demonstrates that the stability of children's responses – like responses obtained from adults – varies to a great extent (0.25 to 1.0). However, no comparisons with adults or within the sample were done. By contrast, Marsh (1986, p. 37) differentiated several groups within a sample of children. It was demonstrated that "younger children and children with poorer reading skills are less able to respond appropriately to negative items and that this effect biases the interpretation of their responses" (see Hershey and Hill 1976 for similar results). Amato and Ochiltree (1987) compared a group of 8- and 9-year-old children with a group of 15- and 17-year-old adolescents. They came to the conclusion that data quality is significantly lower in the case of the younger group.

In a study conducted by Borgers and colleagues (2000), years of education (as a proxy indicator for the cognitive–developmental stage) influenced the internal consistency in the case of a large number of multi-item scales. This supports the hypothesis that data quality increases with cognitive growth. In addition, poor reading skills had a negative influence on data quality (Borgers et al. 2000). Also, with ambiguous response scales younger

children produced less item nonresponse than older children. This effect is likely to be related to their cognitive–developmental stage (Borgers and Hox 2001). It is assumed that younger children do not recognize the ambiguity of the response scales, thus leading to more, though less reliable, responses. In support of these results DeLeeuw and Otter (1995) found a clear interaction between age and clarity of a question. The negative effect of question ambiguity on data quality (test–retest correlation) was more pronounced in the case of the younger children than in the case of the older ones, who coped better with ambiguous questions. This indicates that in earlier cognitive stages, children encounter difficulties during the question–answer process. In addition to the age effect, Borgers and Hox (2000) found an effect of education and educational achievement: "the longer children took education the more reliable their responses are" (Borgers and Hox 2000, pp. 14f.). In sum, these findings provide some preliminary support for the assumption of a positive effect of the children's cognitive functioning on the quality of the question–answer process.

Research regarding the adult population supports the hypothesis according to which the cognitive abilities of a respondent influence the question–answer process. Following Krosnick (1992), for cognitively less sophisticated respondents extensive memory search and other cognitive activities are more burdensome and thus these respondents are less likely to perform them. However, without those activities, respondents make a less than optimal effort when formulating a response to a survey question and they are likely to show satisficing behavior (Krosnick 1992). Thus, they do not thoroughly evaluate the question text and each response category. Instead, they select an easily accessible and plausible response category in order to get through the question quickly. As a result, satisficing respondents are more prone to response effects.

Further empirical results and theoretical reasoning suggest that reduced cognitive functioning is one important factor negatively correlated to response quality (Schwarz and Knäuper 2000). Elderly respondents with already reduced cognitive functioning lack the presuppositions for a thorough processing of a survey question. In an analysis conducted by Knäuper (1999) the lack of working memory capacity is offered as an explanation of larger response order effects among the elderly than among younger adults. "The less working memory capacity an individual has available, the less he or she will be able to process, evaluate, and elaborate the presented information. Accordingly, respondents low in working memory should be especially susceptible to response effects in surveys" (Knäuper 1999, p. 349).

One might speculate that these theoretical approaches also apply to children who lack full cognitive functioning due to their still-growing cognitive-developmental functions. However, a rigid test of the correlation of cognitive skills and the question–answer processes of children and adolescents is not yet available.

The present article is designed to fill that gap and to demonstrate that children are subject to much the same cognitive processes as adults. However, at the same time, we aim to prove that children and adolescents do not show response effects to the same extent as adults because of their limited cognitive skills. In seven large-scale field experiments, age and school achievement (as proxy indicators for cognitive skills) are used as explanatory factors. For dependent variables, we rely on results from several experiments on response effects (question order effects, response order effects, scale effects, and the effect of

the numeric values associated with response options). We will test whether these well-known effects for the adult population can be reproduced for children and adolescents. Thus, we will contribute to the general line of research on whether or not the general assumptions regarding the question–answer process in the adult population also apply to under-aged respondents. Furthermore, we will analyze the effect of low versus high degrees of cognitive functioning on the size of the response effects. The following hypotheses are tested in this article:

1.1. Question wording

When responding to a survey question the respondent works through several steps of the question–answer process. In a first step the respondent aims to develop a lexical and pragmatic understanding of the meaning of the question. In order to decode the substantive scope of a question, respondents consider the question wording and the provided response categories (if any), but also surrounding information such as previous questions. It has been demonstrated many times that changing the question order within a standardized questionnaire can affect the perceived meaning of a given item (e.g., Sudman et al. 1996). On the basis of our theoretical reasoning, we assume that less developed cognitive skills will lead to a lesser amount of contextual information being considered when interpreting a particular survey item. As a result, question order effects – as an indicator for the consideration of previously presented information while answering the current question – will be less prevalent or even nonexistent for respondents with limited cognitive skills.

1.2. Response scales

When answering a question concerning frequency (e.g., "How many times did you see a doctor during the past three months?"), respondents consider the frequency scale (if provided) as relevant information in the process of generating a response. They consider the response scale as a representation of the distribution of the variable in the population (Schwarz et al. 1985). Respondents who estimate or guess the response to the frequency question are guided by the range of the response categories provided. On the basis of our theoretical approach, respondents with less developed cognitive skills are more likely to estimate the frequency of a certain activity (e.g., doctor visits), because their limited cognitive capacities hinder them from remembering, enumerating, and counting the instances. As a result, they are guided to a greater extent by the response scale and they show larger scale effects.

1.3. Response order

It is well-known (Tourangeau et al. 2000) that respondents do not devote the same amount of attention and consideration to all response categories of a closed-ended question. Instead, in a self-administered survey, response options presented early in a list of categories are evaluated more thoroughly, and thus are selected more often, than ones at the end of the list. In the literature several examples of this so-called primacy effect are reported (Krosnick and Alwin 1987). It is assumed that respondents evaluate the first few categories in depth; however, the respondents' cognitive capacity is not large enough to be

spread proportionally across all categories. Because children are assumed to have limited cognitive skills as compared to adults, they will be even less able to assign the same amount of attention to all response categories, thus showing larger response order effects.

1.4. Numeric values associated with the responses categories

On certain attitude scales, researchers do not provide a full verbal expression for each response option. Instead, they label the end-points of the response scale with verbal expressions and mark response categories in between with numeric values only (Schwarz et al. 1991). This leaves it up to the respondent to infer the meaning of the response categories on their own, which is a demanding task that cannot be adequately performed without a certain degree of intellectual skill. In line with the theoretical reasoning discussed above, respondents with less developed cognitive abilities have difficulty in decoding the meaning of the response categories between the labeled end-points. As a consequence, they are more likely to rely on the numeric values provided with the unlabeled response categories – which are easier than decoding the response scale on the basis of spare verbal expressions at the end-points of the scale. When the range of the numeric values is changed, respondents will be guided differently. Because respondents with less developed cognitive skills more often rely on the numeric values, they should be affected by modifications of the numeric values to a greater extent.

2. Methods

In this article we will report results from field experiments originally designed to investigate the effect of child characteristics on response error. The experiments were embedded in three large-scale self-administered surveys conducted between 1999 and 2001 with children, adolescents, and young adults aged 10 to 21 ($n = 4{,}206$; $n = 5{,}042$; $n = 1{,}200$). Each survey had its own substantive topic and goal.

2.1. Study 1

Study 1 was conducted in the spring of 1999 as a second wave of an ongoing longitudinal study on violence in schools in Bavaria (a state in Germany). 4,206 respondents aged 10 through 21 (some older) completed a self-administered paper questionnaire in a classroom setting (see Fuchs et al. 2001 for details on the study). A cluster sample stratified by educational track was used in order to survey a representative sample of students in Bavaria.

2.2. Study 2

Study 2 was fielded in the spring of 2001 as a self-administered survey on right-wing attitudes (see Fuchs et al. 2003 for details). 5,042 children, adolescents and young adults aged 13 through 21 (some older) completed a questionnaire in class (three quarters of the sample completed a paper questionnaire, one quarter completed an online survey, also in class). The sample design applied a clustered design stratified by educational track and yielded a representative sample of students for the state of Bavaria.

2.3. Study 3

Study 3 focused on conflicts between generations. It was administered in the spring of 2000 as a mail survey based on a register sample. 1,200 respondents stratified by several age groups participated in the survey (adolescents, adults, elderly). This study allows a comparison of adolescent respondents aged 15 through 20, adults aged 21 through 60, and the elderly aged 61 through 90.

Several experiments on question order effects, response order effects, scale effects and the effect of the numeric values associated with response categories were incorporated in the three studies. All experiments were designed as split experiments (between subject design) with the respondents having been randomly assigned to one out of two or four split groups, depending on the design of the particular experiment (see Table 1 for details of the experiment). Each group received a carefully manipulated questionnaire. Comparing the response distributions of the experimental variables across the split groups, we were able to compute the size of the response effects. Due to the large number of cases, we could also calculate the size of the response effects for subgroups of respondents with low vs. high degrees of cognitive skills. A comparison of the size of the response effects for these subgroups provided us with data to test our hypotheses.

We were not able to measure cognitive skills directly. Instead, we differentiated our sample by age groups and school achievement. With respect to the age distribution in each study, we have collapsed respondents within groups of an age range of 2 to 4 years; in Study 3, we had to build larger groups due to the age range of the subjects and also due to the smaller sample size. Unfortunately, this makes comparisons across experiments difficult. Also, the age group breakdowns do not correspond with the definition of children and adolescents in official statistics or with those in the law. In Studies 1 and 2, school achievement was measured using the student's self-report regarding the teacher's satisfaction with the student's performances in school. Of the five-point response categories, we collapsed the response categories "very satisfied" and "satisfied" into "good school achievement;" all other categories ("partly satisfied," "not satisfied," "not satisfied at all") were collapsed into "intermediate/poor school achievement." In Study 3, we used

Table 1. Overview of the experiments in the three studies

Experiments	Study #	Split versions	*N*
Response order effect			
Experiment 1	1	4	4,183
Experiment 2	2	2	2,017
Scale effect			
Experiment 3	2	2	4,686
Experiment 4	3	2	817
Effect of numeric values			
Experiment 5	2	2	4,791
Question order effect			
Experiment 6	2	2	4,813
Experiment 7	2	2	4,699

N indicates the number of cases considered for each particular experiment (excluding item missing values and cases that were not included due to skip patterns in the questionnaire).

the highest educational degree (for adults) and the aspired-to educational degree (for students). Germany has a tracked educational system with three major degrees and many minor variations. For the purpose of our analysis we have collapsed the highest degree and some of its variations ("high degree") into one group and compared it to the middle and lower degrees ("middle/low degree").

We have to admit that our indicators for cognitive skills are based on an indirect assessment. Ideally, we would rather have administered psychological tests on working memory and other cognitive skills in order to obtain a valid assessment of the respondents' cognitive capacities. However, such a detailed measurement would have required more time in the interview situation and space in the questionnaire, neither of which was available.

3. Results

The following sections summarize results from the field experiments. First we focus on response order effects (two experiments, Section 3.1). Next we discuss scale effects (two experiments, Section 3.2) and the effects of the numeric values associated with the response categories (one experiment, Section 3.3). In the final section we summarize the results regarding question order effects (two experiments, Section 3.4).

3.1. Response order effects

3.1.1. Experiment 1

In Experiment 1 (as part of Study 1), a long list of response options (12 categories) was presented to the children and adolescents. The question focused on the infrastructure of the respondents' neighborhoods. Respondents were supposed to mark all response options that apply. The response options were presented to the respondents in four different versions: The original version, a reversed version (see Figure 1) and two other versions in which the response options positioned first in the original version were presented in the second and third quarters of the list.

31. Which of the following facilities are available in your neighborhood? (mark all that applies)	31. Which of the following facilities are available in your neighborhood? (mark all that applies)
☐ Movie theater	☐ Grocery store
☐ Youth center	☐ Playground
☐ Indoor pool	☐ Preschool
☐ School	☐ Museum
☐ Bus stop	☐ Theater
☐ Library	☐ Library
☐ Theater	☐ Bus stop
☐ Museum	☐ School
☐ Preschool	☐ Indoor pool
☐ Playground	☐ Youth center
☐ Grocery store	☐ Movie theater

Fig. 1. Experimental questionnaire versions of Experiment 1, response order effect, original version (left) and reversed version (right), other versions not shown

In this analysis results are presented for the response option "movie theater" (= a movie theater is present in the respondent's neighborhood). As expected, on the basis of the findings in the literature for self-administered surveys (Krosnick and Alwin 1987; Schwarz et al. 1992), this experiment reveals a moderate primacy effect. When "movie theater" is offered at the beginning of the list, a significantly larger proportion of respondents report a movie theater in their neighborhood (5 percentage points; $p < 0.001$) as compared to the respondents confronted with the other versions (with "movie theater" in the second or third quarter of the list of response categories or at its end).

As shown in Table 2, the response order effect correlates with age: younger respondents aged 10 through 13 show a larger effect (10 percentage points) than older respondents (3 or 4 percentage points $p < 0.01$). In addition, school achievement is also correlated ($p < 0.05$): pupils with higher ratings of their school achievement show smaller response order effects (4 percentage points) than pupils with intermediate or poor school achievement (6 percentage points).

Table 2. Size of response order effect (primacy effect, percentage points) by age and school achievement

	School achievement		
Age groups	Good	Intermediate/Poor	Total
10–13	7*	11*	10***
14–17	1	5	3
18 and older	5	2	4
Total	4*	6**	5***

Study #1, Question #31, Category "movie theater."
Displayed is the percentage point difference of the endorsement of "movie theater" being presented at the beginning of the list (original order) minus the proportion of respondents selecting the same response option when it is presented in the second, third or fourth quarter of the list (reversed version and other versions). The response option "grocery store" could have been analyzed as well; however, the response order effect is considerably smaller and does not reach statistical significance.
*< 0.05, **$p < 0.01$, ***$p < 0.001$ for the effect itself; test for differences of the size of the response effects: age $p < 0.01$, school achievement $p < 0.05$.
Note: The level of significance given for the values in the table indicates the results of a chi^2-test for the effect itself (= test on equal distributions of the experimental variables across the two questionnaire versions). In order to test the differences in the size of the response effects across the age groups or the levels of school achievement, we apply a separate test: due to the fact that we are comparing differences of percentage points across independent samples, no standard test of significance is available. For the purposes of this analysis we have computed a test of the observed distribution of the response effects against the assumption of equal sizes of the response effects for all groups.

On the basis of our theoretical reasoning we assumed that the limited cognitive skills of children hinder them to a higher degree from responding adequately to our survey questions as compared to adolescents and adult respondents. Even though we did not measure cognitive skills directly, we may interpret these findings in support of our hypothesis: younger respondents and respondents with less than optimal school achievement do not process all responses equally comprehensively, and thus they show larger response order effects.

3.1.2. Experiment 2

In Experiment 2 (incorporated in Study 2) two versions of a question with a long list of response options (ten response categories) were randomly assigned to the subjects.

Respondents were supposed to select one response category that best describes the youth style or youth culture of which they consider themselves members. The second version contained the same response options as the original version but in reversed order (see Figure 2).

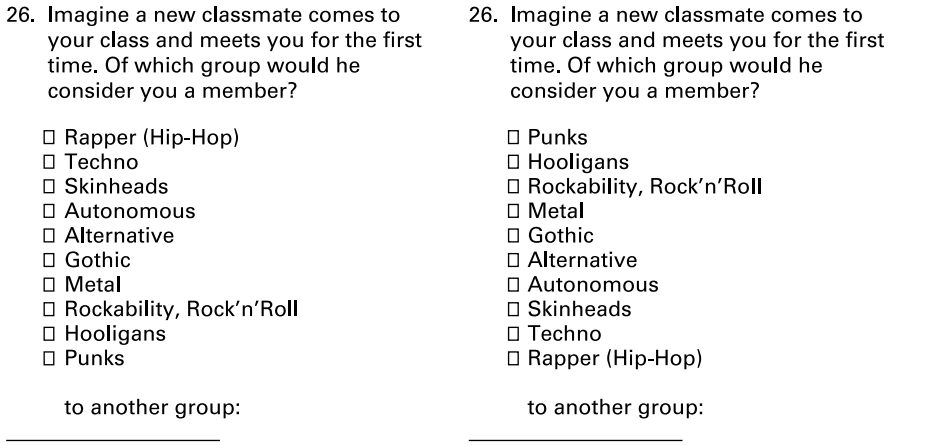

Fig. 2. Experimental questionnaire versions of Experiment 2, original order of response categories (left) and reversed order (right)

On the basis of the findings published in the literature (Krosnick and Alwin 1987), primacy effects were to be expected: in a self-administered survey mode, response categories presented first in the list of options receive more attention and they establish an interpretative framework or standard of comparison for subsequent options. In addition, "a satisficing criterion implies that the earlier in the list an acceptable answer appears, the more popular it will be. If respondents choose an early option, they can avoid thinking about the others" (Tourangeau et al. 2000, p. 250).

Overall we observe a response order effect of 6 percentage points ($p < 0.001$). To our surprise and against the common wisdom in the literature (e.g., Hippler and Schwarz 1992), no primacy effect but rather a recency effect appears in the data. The first two response categories in the original version are endorsed by a higher proportion of respondents if presented at the end of the list in the reversed questionnaire version. Even though we do not have a compelling explanation for this effect we attribute it to the fact that below the list of response options a line with the label "to another group" was offered. Presumably the respondent's focus jumped to this line and then went through the list of response alternatives from bottom up. Another plausible explanation is based on the assumption that a primacy effect is expected for a plausible item but a recency effect is expected for an implausible one. Looking at the content of the question, one might come to the conclusion that many of the group identifications for young people are likely to be implausible.

However, our initial goal was to study response order effects in the presence of different cognitive skills, thus we do not go into greater detail regarding the reasons for the appearance of a recency effect. Instead, we further differentiate the sample into groups of high vs low cognitive functioning on the basis of age and school achievement.

Considering the theoretical reasoning summarized in the first section, we expected older respondents and respondents with better school achievement to be less prone to response order effects. They are in a better position to hold all response options in working memory at a time, and thus they are able to spread their attention more equally to all response categories. This should result in less pronounced response order effects for the older respondents and respondents with better school achievement. By contrast, younger respondents and students with intermediate or poor school achievement are expected to show larger response order effects.

As expected, the size of the effect described above depends on age (see Table 3): younger respondents show larger effects than adolescents (age 16 through 17) and adult respondents ($p < 0.07$). Even though the recency effect by itself is contrary to expectations, the age effect is as predicted. By contrast, school achievement has no significant effect on the size of the response order effect for the older respondents. However, for the respondents aged 13 through 15 we found an anomalous significant difference between the sizes of the response effects that is contrary to our predictions: the response order effect is larger for respondents with good school achievement than for ones with intermediate and poor school achievement.

Table 3. Size of response order effect (recency effect, percentage points) by age and school achievement

	School achievement		
Age groups	Good	Intermediate/Poor	Total
13–15	19*	10*	11***
16–17	10	2	4
18 and older	−1	0	−1
Total	7	5	6***

Study #2, Question #26, Categories 9 and 10.
Displayed is the percentage point difference of the combined endorsement of the first two response categories being presented at the end of the list (reversed order) minus the combined proportion of respondents selecting the same response options being presented at the beginning of the list (original order). The option "to another group" was not considered in this computation. Also, the results for response options presented at the beginning of the list in the original order are not displayed – they are smaller in size.
$*p < 0.05$, $***p < 0.001$ for the effect itself; test for differences of the size of the response effects: age $p < 0.07$, school achievement not significant; see note to Table 2 for remarks on statistical testing.

3.2. Scale effects

In addition to the experiments on response order effects, we conducted several experiments on scale effects, that is, the effect of the range of the response options provided in a frequency question. This experiment is based on a classic design described by Schwarz and colleagues (1985), where respondents were asked to report their daily TV consumption on a frequency scale. A random subset of the respondents was provided with response categories ranging from "up to $\frac{1}{2}$ hour" through "more than $2\frac{1}{2}$ hours". The other half of the respondents received response categories ranging from "up to $2\frac{1}{2}$ hours" through "more than $4\frac{1}{2}$ hours".

The design of the experiment is based on the following theoretical reasoning. Some people might be able to report their daily TV consumption without some kind of

estimation. However, most respondents need to initiate a complex estimation process in order to produce an adequate response to the question. Instead of recalling a stored response they will presumably estimate the duration based on the response scale provided. It is assumed that respondents consider the range of the response options as a representation of the distribution of the variable (TV consumption) in the population, where the midpoint of the scale represents the population average and the extremes of the scale indicate the high and the low TV consumption in the population, respectively. Their own response will be formatted according to the response scale printed in the questionnaire, e.g., a response like "well, I watch more than most people, but I am not extreme" might be formatted into the second highest response category. Because they have no detailed knowledge regarding the duration of their daily TV consumption they do not evaluate the verbal label associated with the second highest response option in greater detail (e.g., "up to $4\frac{1}{2}$ hours"). Accordingly, in the study conducted by Schwarz and colleagues (1985) respondents confronted with the high-frequency scale reported a significantly longer daily TV consumption than respondents offered the low-frequency scale.

3.2.1. Experiment 3

In Experiment 3 we replicated this experiment with children and adolescents (incorporated in Study 2; see Figure 3). Respondents were asked to report their daily TV consumption on a 6-point scale according to the original response alternatives described by Schwarz and colleagues (1985). On the basis of the assumptions discussed above, respondents are expected to report a longer daily TV consumption when provided with a high-frequency scale than when provided with a low-frequency one (= scale effect).

Overall we observe a pronounced scale effect (see Table 4): the proportion of respondents choosing a response option indicating a TV consumption of more than $2\frac{1}{2}$ hours a day is 25 percentage points higher ($p < 0.001$) when using the high-frequency response scale than when using the low-frequency one. These findings are pretty much in accordance with the results reported in the literature (Schwarz et al. 1985).

On the basis of the theoretical approach described in the introduction, respondents with less developed cognitive skills are more likely to estimate their TV consumption, because their limited cognitive capacities do not allow them to remember all instances of watching TV, to add the durations and finally to compute an average based on several days. As a result they try to estimate or guess a response and while doing so they are guided to a

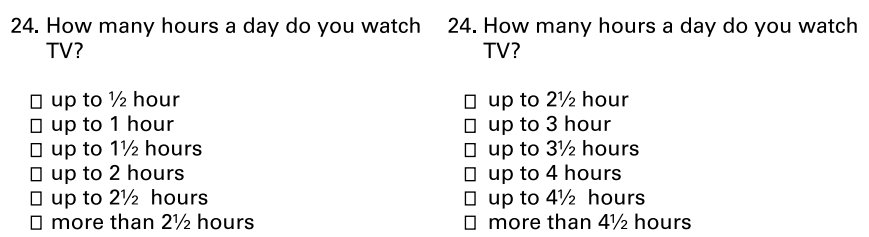

Fig. 3. Experimental questionnaire versions of Experiment 3, scale effect "TV consumption", low-frequency scale (left), high-frequency scale (right)

Table 4. Size of scale effect (percentage points) by age and school achievement

	School achievement		
Age groups	Good	Intermediate/Poor	Total
13–15	25***	32***	30***
16–17	19***	28***	25***
18 and older	22***	17***	19***
Total	22***	27***	25***

Study #2, Question #24.
Displayed is the percentage point difference of the endorsement of a response option indicating TV consumption of more than $2\frac{1}{2}$ hours a day on the high-frequency scale minus the corresponding proportion on the low-frequency scale.
*** $p < 0.001$ for the effect itself; test for differences of the size of the response effects: age $p < 0.01$, school achievement $p < 0.05$; see note to Table 2 for remarks on statistical testing.

greater extent by the response scale provided in the questionnaire, thus showing larger scale effects.

As expected, younger respondents show larger scale effects. While the scale effect reaches 19 percentage points for respondents aged 18 and up, adolescents aged 16 and 17 show an effect of 25 percentage points. Consistent with this increase, the effect reaches 30 percentage points in the youngest group ($p < 0.01$). Also, respondents with poor and intermediate school achievement show an effect of 27 percentage points, as compared to 22 percentage points in the subgroup with good school achievement ($p < 0.05$).

Both findings support our hypothesis according to which better cognitive functioning allows respondents to generate a response more independently from the response categories provided in the questionnaire. They are in a better position to determine their daily TV consumption on the basis of internal information, which is a demanding task. By contrast, respondents with less developed cognitive skills rely more heavily on questionnaire characteristics alone (e.g., the response scale) when estimating their daily TV consumption.

3.2.2. Experiment 4

In a second experiment on scale effects the same classic design described by Schwarz and colleagues (1985) was applied. Respondents in Study 3 were asked to report their daily TV consumption on a six-point scale. Again, two experimental versions of the response scale were presented to two random halves of the sample (see Experiment 3 and Figure 3 for details of the design of the high-frequency vs low-frequency scales). As described before, respondents confronted with the high-frequency scale were expected to report a longer daily TV consumption than those confronted with the low-frequency scale. The results confirm this prediction (Table 5). Overall, the proportion of respondents reporting a daily TV consumption of more than $2\frac{1}{2}$ hours is 31 percentage points higher when using the high-frequency scale than when using the low-frequency scale ($p < 0.001$).

Looking at the size of the scale effects one can easily detect pronounced educational differences ($p < 0.001$): respondents in the lower or intermediate educational track (or having completed a lower or intermediate degree) show a significantly larger scale effect (33 percentage points) than the respondents with a higher degree (22 percentage points,

Table 5. Size of scale effect (percentage points) by age and school track

Age groups	School track/degree		
	Higher track/degree	Intermediate/Lower track/degree	Total
15–20	24^{+}	37***	34***
21–60	20**	25**	26***
61–90	21^{+}	41***	35***
Total	22***	33***	31***

Study #3, Question #6–2.
Displayed is the percentage point difference of the endorsement of a response option indicating TV consumption of more than $2\frac{1}{2}$ hours a day on the high-frequency scale minus the corresponding proportion on the low-frequency scale.
$^{+}< 0.10$, $**p < 0.01$, $***p < 0.001$ for the effect itself; test for differences in the size of the response effects: age not significant, educational level $p < 0.001$; see note to Table 2 for remarks on statistical testing.

$p < 0.001$). This is especially true for respondents aged 15 through 20 and also for respondents aged 61 and older.

In addition, for the younger groups of respondents (34 percentage points) as well as for the older group (35 percentage points) the scale effects are substantially larger than for the adult population aged 21 to 60 (26 percentage points, not statistically significant). Again, we have to speculate what the reasons may be. At this point it is assumed that respondents with poor cognitive skills have greater difficulty answering the question based on hard-to-process internal information (information regarding the actual event of watching TV). Instead, they rely on easily accessible questionnaire characteristics (the response scale) to a greater extent. By contrast, for respondents with better-developed cognitive skills the processing of the necessary internal information is a less demanding and maybe even a more enjoyable task. This could be interpreted as a hint towards the hypothesis that young and old respondents suffer from a lack of cognitive resources to a similar extent, and that this circumstance might be responsible for the more pronounced magnitude of the response effects in both groups. However, we have to be careful when generalizing findings and explanations for response effects among the elderly (e.g., Knäuper 1999) to under-aged respondents.

3.3. Numeric values associated with response categories

3.3.1. Experiment 5

In another experiment (incorporated in Study 2) we assessed the effect of different numeric values associated with response options in an attitude question. We replicated an experiment published by Schwarz and colleagues (1991) where respondents were asked to rate their "overall success in life" on an 11-point scale. The end-points of the scale were labeled "not at all successful" and "very successful," respectively. The remaining scale points were not verbally labeled but were marked by numeric values. Two different versions of the question were presented to random subsets of the sample. One half received a questionnaire with numeric values ranging from -5 through $+5$ and the other half of the sample answered the same question with numeric values from 0 through 10 (see Figure 4).

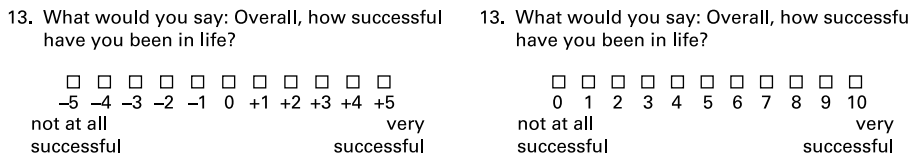

Fig. 4. Experimental questionnaire versions of Experiment 5, effects of numeric values associated with response options, "overall success in life," −5/+5 version (left), 0/10 version (right)

We expected respondents to encounter difficulty when decoding the meaning of the underlying evaluative dimension of the response categories when text labels were missing. Thus, they would have to rely on the numeric values associated with the response options. On the basis of the findings reported in the literature (Schwarz et al. 1991b) a significant shift in the overall rating of the success in life is to be expected depending on the range of the numeric values. Respondents confronted with values ranging from −5 through +5 would presumably interpret the value 0 as the neutral midpoint of the scale ("not successful") while −5 would be interpreted as the opposite of success, i.e., failure (Schwarz and Hipper 1995). By contrast, respondents confronted with the same response options labeled 0 through 10 were expected to interpret the left extreme point of the scale as intended by the researcher ("not successful at all"). Thus, the overall ratings of the respondents using the scale labeled −5 through +5 should be shifted towards the right end-point, that is, towards the "very successful" extreme.

As expected, we observe a considerable shift of the response distribution depending on the numeric values associated with the scale points (see Table 6). When confronted with the scale ranging from −5 through +5, the proportion of respondents choosing a response above the midpoint is 13 percentage points higher ($p < 0.001$) as compared to the 0/10 version. This result confirms our expectations and follows the patterns described in the literature (Schwarz et al. 1991b; Schwarz and Hippler 1995).

Again we use the size of the resulting response effect as a dependent variable in order to assess the impact of age and school achievement on the question–answer process. Assuming that younger children and children with intermediate or poor school

Table 6. Effect of numeric values associated with response categories (percentage points) by age and school achievement

	School achievement		
Age groups	Good	Intermediate/Poor	Total
13–15	9*	18***	16***
16–17	8*	13***	12***
18 and older	8**	12***	11***
Total	9***	15***	13***

Study #2, Question #13.

Displayed is the percentage point difference of the proportion of students choosing a response category above the scale mid-point in the −5/+5 version minus the proportion of students choosing a response category above mid-point in the 0/10 version.

*< 0.05, ** $p < 0.01$,*** $p < 0.001$ for the effect itself; test for differences in the size of the response effects: age $p < 0.01$, school achievement $p < 0.001$; see note to Table 2 for remarks on statistical testing.

achievement are characterized by less developed cognitive skills, we predicted an effect of these variables on the magnitude of the response effect. Cognitively less skilled respondents are less able to deal with the complicated task of building a mental representation of the evaluative dimension of the question based on the verbal labels provided at the end-points of the scale. Hence, they make use of the easily accessible numeric values associated with the response categories to a greater extent. Thus, the response effect will be larger for the younger and less sophisticated group.

As expected, older respondents show a less pronounced response effect (11 percentage points) than younger respondents (16 percentage points, $p < 0.01$). In addition, we observe a pronounced difference comparing respondents with poor or intermediate school grades (15 percentage points) vs good school achievement (9 percentage points, $p < 0.001$).

3.4. Question order effects

In this final section we focus on question order effects. So far, less developed cognitive skills contribute to more pronounced response order effects, scale effects, and effects of the numeric values associated with response categories. By contrast, for question order effects we predict a contrary correlation. The less developed the cognitive skills are, the less pronounced the question order effects should be. In order to recognize the preceding questions as providing a relevant context for the understanding of the current question, respondents need to store the information provided in the previous question in short-term memory. In addition, they need to link the semantic and pragmatic meaning of the previous question to that of the current one. Both tasks are more demanding for respondents with limited cognitive abilities. Thus, this group will show less pronounced question order effects.

3.4.1. Experiment 6

In the first experiment on question order (incorporated in Study 2), respondents were asked to rate their chances of finding a job after school as compared to those of other adolescents in the eastern or western part of Germany. Two questions were asked: one comparing the respondent's chances with those of the average juvenile in western Germany, the other comparing the respondent's chances with those of the average juvenile in eastern Germany (all respondents were living in the western part of Germany). The two questions were asked in the original order (west first) or in the reversed order (east first). Respondents were randomly assigned to either the original or the reversed question order (see Figure 5).

We expected the question order to have an effect on the perceived meaning of the questions. Asking the respondents to compare themselves to the juveniles in the eastern part first (before being asked the question regarding the western juveniles) should increase the individual's perception of his or her chances as compared to those of the average juvenile in the western part of Germany. Even though a respondent might see little or no chance at all compared to the other juveniles in the western part of Germany, he or she might arrive at a less negative perception regarding this comparison if asked about the juveniles in the economically less prosperous eastern part of Germany first (e.g., "Compared to them, my chances in the western part of Germany are not that bad at all!").

11. Compared to other juveniles *in the western part of Germany,* in your own perspective do you have better chances of finding a job, worse chances or about the same chances?

☐ better chances
☐ about the same chances
☐ worse chances

12. Compared to other juveniles *in the eastern part of Germany,* in your own perspective do you have better chances of finding a job, worse chances or about the same chances?

☐ better chances
☐ about the same chances
☐ worse chances

11. Compared to other juveniles *in the eastern part of Germany,* in your own perspective do you have better chances of finding a job, worse chances or about the same chances?

☐ better chances
☐ about the same chances
☐ worse chances

12. Compared to other juveniles *in the western part of Germany,* in your own perspective do you have better chances of finding a job, worse chances or about the same chances?

☐ better chances
☐ about the same chances
☐ worse chances

Fig. 5. Experimental questionnaire versions of Experiment 6, question order effect, "chances of finding a job after school," original order (left), reversed order (right)

Overall, we find a moderate question order effect of 6 percentage points in the expected direction (see Table 7). Respondents who are confronted with the reversed question order rate their chances of finding a job as significantly better ($p < 0.01$) than respondents answering the same question in the original order.

As predicted, our data show a relationship of the size of the question order effect with age ($p < 0.01$). For older respondents we observe substantial effects while younger respondents do not show response differences. This result may be interpreted in support of the hypothesis. Young respondents lack cognitive skills. Thus, they do not recognize a preceding question as providing a relevant context for the understanding and processing of the current question. In addition, respondents with good school achievement show larger question order effects than the respondents with poor or intermediate school achievement. Even though the latter difference does not reach statistical significance ($p < 0.11$), it is also in line with our hypothesis.

Table 7. Size of question order effect, "chances of finding a job" (percentage points) by age and school achievement

	School achievement		
Age groups	Good	Intermediate/Poor	Total
13–15	2	0	0
16–17	7	−1	1
18 and older	13	6	10
Total	7**	5**	6**

Study #2, Question #11.
Displayed is the percentage point difference of the proportion of students expecting "better chances of finding a job compared to other juveniles *in the western part of Germany*" in the reversed version minus the corresponding proportion in the original order.
**$p < 0.01$ for the effect itself; test for differences in the size of the response effects: age $p < 0.01$, school achievement $p < 0.11$; see note to Table 2 for remarks on statistical testing.

3.4.2. Experiment 7

In a second experiment on question order effects (incorporated in Study 2) respondents were asked to rate their satisfaction with their life in general and also their satisfaction with their school class. This experiment applied a design adopted from Schwarz's classic experiment concerning satisfaction with life and marriage (Schwarz et al. 1991a). Random subsets of respondents were presented with two versions of the questionnaire (see Figure 6). Version 1 asked the question about satisfaction with life in general first and then the question regarding satisfaction with their class. Version 2 asked the same two questions in reversed order. Owing to the respondents' interpreting the question about their class differently when asked about life in general first, we expected a question order effect: when asked about satisfaction with life first, by contrast satisfaction with their class receives a considerably lower rating. Under this condition their class is seen as a place of learning and examinations but not as a place where one can meet friends and socialize (like life in general). When asked about their satisfaction with their class first, the question conjures up a wider image to the respondents, including friends and activities during breaks, because these areas were – unlike in the original version – not covered by a previous question. Hence, their class receives a better rating when asked about first. (It should be noted that the classic marriage–life experiment yields a question order effect for the general question on life, not for the specific one on marriage.)

On average we found a moderate question order effect (7 percentage points; $p < 0.001$; see Table 8). Respondents being asked the question regarding their class first end up with a significantly better rating of satisfaction with their class as compared to those who are asked about their satisfaction with life in general first.

This question order effect was also differentiated by age and school achievement. The result of this analysis indicates smaller question order effects for the younger respondents than for the older ones ($p < 0.001$). Also, for children and adolescents with good school achievement we compute slightly larger question order effects (not statistically significant) than for ones with poor or intermediate school achievement.

Taking the findings of the two experiments on question order together, the results are generally in support of the hypothesis according to which respondents with better cognitive functioning are more likely to treat preceding questions as a relevant context for

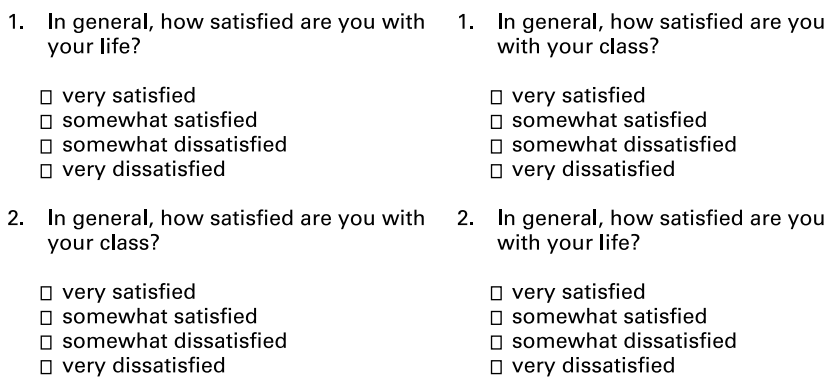

Fig. 6. Experimental questionnaire versions of Experiment 7, question order effect, "satisfaction with life/class" original order (left), reversed order (right)

Table 8. Size of question order effect, "satisfaction with class" (percentage points) by age and school achievement

	School achievement		
Age groups	Good	Intermediate/Poor	Total
13–15	9*	4*	5*
16–17	6^{+}	8***	7***
18 and older	11***	7*	9***
Total	8***	7***	7***

Study #2, Question #2 (This question was not applicable to vocational school students who go to school one day a week in addition to their job or apprenticeship; they were excluded from the analysis.)
Displayed is the percentage point difference of the proportion of students who endorse "very satisfied" or "somewhat satisfied" in the reversed version minus the corresponding proportions in the original order.
$^{+} < 0.10$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$ for the effect itself; test for differences of the size of the response effects: age $p < 0.001$, school achievement not significant; see note to Table 2 for remarks on statistical testing.

the understanding and processing of the current question. So far we were able to reach the level of statistical significance for the differences in the size of the response effects by age groups. By contrast, the differences by school achievement missed significance in Experiment 6 as well as in Experiment 7.

4. Summary and Discussion

The results presented in this article are relevant to the general question to what extent the available methodological knowledge regarding the question–answer process applies also to children and adolescents. On the basis of the assumption that children lack full cognitive functioning because of their cognitive-developmental stage, we predicted a less sophisticated question–answer process for these respondents. In order to assess this hypothesis we have focused on age and educational achievement as explanatory factors, while the size of response effects was treated as an indicator for the cognitive processes in the respondents' minds when responding to a survey question. In order to assess multiple aspects of the question–answer process, four types of experiments were conducted: question order experiments in order to assess the respondents' understanding of a question, experiments on response order effects in order to assess the retrieval of a response, and finally, experiments on the effects of the numeric values associated with the response options and on scale effects in order to assess the formatting stages in the question–answer process.

Looking at the results in general (see Table 9), every difference in the size of the response effects by age groups is in support of our hypothesis. With the exception of Experiment 4, these differences reach statistical significance. In sum, the size of response order effects as well as that of scale effects and the effects of numeric values associated with response categories decreases with age. In general, for each age group comparison, the response effects mentioned above reach the highest level among the youngest group, which is assumed to be cognitively less developed. At the same time, younger respondents are more likely to ignore contextual information when decoding the meaning than are older respondents.

Table 9. Summary of results, "Are there any differences in the magnitude of the response effects in support of the hypothesis? Yes/no," significance of difference

	Indicators for cognitive skills	
	Age	School achievement, education
Response order effect		
Experiment 1	Yes**	Yes*
Experiment 2	Yes^{+}	No
Scale effect		
Experiment 3	Yes**	Yes*
Experiment 4	Yes	Yes***
Effect of numeric values		
Experiment 5	Yes**	Yes***
Question order effect		
Experiment 6	Yes**	Yes
Experiment 7	Yes**	Yes

$^{+}p < 0.10$; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

With some restrictions, this holds also for the differences in the size of the response effects for children and adolescents with good vs intermediate/poor school achievement. The only thing is, that the results of Experiment 1 do not support the assumption that less sophisticated respondents show a less sophisticated question–answer process. By contrast, most of the other question order effects, scale effects, effect of numeric values and also one response order effect are in support of the hypothesis. With the exception of the Experiments 6 and 7 on question order, the differences are statistically significant.

In general, the results support the hypothesis according to which children with well-developed cognitive skills answer survey questions similarly to adults while younger children and adolescents with poor school performance lack the cognitive capacity to do so. Thus, they treat the questions as more separated and isolated from each other (smaller question order effects) and rely more often on the response scale. In addition, they are guided to a greater extent by the response order or the numeric values associated with the response options.

So far we do not have a complete picture. However, some basic thoughts should be considered when designing questions for young respondents. Children up to approximately age 14 still have a limited understanding of the questions. Nevertheless, they provide us with a response because they are guided by the question text and response alternatives to a larger extent than older respondents. To emphasize it, two processes are counter-compensating for each other. Limited cognitive skills of young respondents lead to a less complete understanding of a question. At the same time, because of their less developed cognitive skills, they also rely more heavily on the information provided in the questionnaire in order to produce an answer. Thus, on the surface, data obtained from young respondents look similar to those received from older respondents, and item nonresponse is not a big issue in surveys with young children (Borgers and Hox 2001; Fuchs 2003). However, the former are based on a less sophisticated question–answer process, because children and adolescents respond even to complicated questions because

their limited cognitive skills preserve them from recognizing that they should have a problem with these items. When they become older, they get a more complete picture of the meaning; however, at the same time they are in a better position to reliably respond to the question and they are decreasingly guided by the information provided in the questionnaire.

The research on children and adolescents as respondents is still in its infancy. Thus, some methodological aspects for future research should be emphasized. In this article, age and school achievement are used as proxy indicators for cognitive skills. However, we are well aware that the human cognitive-developmental growth does not follow a strict age-related logic, and that for this reason some children may reach a certain cognitive stage earlier than others. School achievement may instead be better interpreted as an indicator of the individual's performance, of motivation and the like. However, many of the experiments reported in this article yield an interaction effect of age and educational achievement and while they both seem to contribute to response effects in the same direction, we are not sure what exactly they stand for. Future research should take advantage of the available psychological tools and scales in order to assess the dimensions of cognitive functioning in greater detail.

Furthermore, additional indicators for the question–answer process in the respondent's mind should be considered for assessment. The indicators used in the present article are in a way still distant from the cognitive processes themselves. In order to develop more detailed indicators it might be worth doing cognitive interviewing with children. Previous research has demonstrated that cognitive interviewing is a suitable technique to asses the question–answer process in the case of children (Zukerberg and Hess 1997; Hess et al. 1998). Presumably, this would increase our knowledge regarding their question understanding, answer retrieval, and formatting strategies.

The question–answer process is a complex interaction of the respondent with the questionnaire. In this article we have assessed the effect of the cognitive skills of children and adolescents on this process. So far, we have not evaluated mode differences and their effect on the question–answer process in samples of children and adolescents. Previous research was able to demonstrate (Lyberg and Kasprzyk 1991; DeLeeuw 1993; Holbrook et al. 2003) that differences introduced by the specifics of the applied mode of administration might affect data quality. So far, our experiments rely on data obtained through self-administered surveys and it is unknown whether the effects described in this article would also appear in a face-to-face or telephone interview.

As mentioned in the methods section, the age-group breakdowns used in the analyses make it difficult to compare results across experiments. In addition, the age groups used do not correspond with the definition of children and adolescents in official statistics or in the law. In a next step, we are looking for a large-scale survey with children and juveniles in order to administer some of the experiments reported in this article to a larger sample of children and juveniles. This would provide us with the data to compute all response effects for the same "official" age groups. While the results of the experiments reported in this article demonstrate a general impact of cognitive functioning of children and adolescents on the quality of the question–answer process, a large sample of children and adolescents in standardized age groups would allow us to model the increase or decrease of response effects in dependence upon the development of cognitive functioning.

We should also admit that Studies 1 and 2 were administered in a classroom setting, which implies additional restrictions for generalizations. The classroom atmosphere might have reduced the respondents' perception of the survey as a specific communicational and social phenomenon and they might have misinterpreted it as an examination. Hershey and Hill came to a similar conclusion according to which "second and fourth graders frequently showed less tendency towards response set than older respondents did" (1976, pp. 713f.). The authors attribute this finding to the younger children's "test-taking mentality" and their more serious attitude towards the survey. So far, it is not clear what consequences this attitude might have had with regard to data quality. One might speculate that response effects are in general underestimated under classroom conditions, especially in the case of the younger respondents. On the basis of this interpretation we would be on the safe side because the difference age makes regarding the size of the response effects should be even bigger in real life. However, at the same time, based on the assumption of a "test-taking mentality" one could expect different degrees of motivation among students depending on the level of school achievement. Accordingly, we have to be especially careful in attributing the differences between good and intermediate/poor school performers to their different levels of cognitive skills. Hence, future research should try to replicate the finding under other mode conditions.

Previous research has demonstrated that questions differ in terms of formal characteristics as well as in terms of the difficulty in answering them (Borgers et al. 2003a, 2003b). The results reported in this article rely on experiments with questions relatively easy to answer. Therefore, it would be of great value to replicate the experiments with questions of different types and difficulty (formal difficulty, e.g., number of response options provided, as well as substantive difficulty, e.g., questions on sensitive issues).

5. References

Amato, P. and Ochiltree, G. (1987). Interviewing Children About Their Families: A Note on Data Quality. Journal of Marriage and the Family, 49, 669–675.

Benson, J. and Hocevar, D. (1985). The Impact of Item Phrasing on the Validity of Attitude Scales for Elementary School Children. Journal of Educational Measurement, 22, 231–240.

Borgers, N., DeLeeuw, E., and Hox, J.P. (2000). Children as Respondents in Survey Research: Cognitive Development and Response Quality. Bulletin de Methodologie Sociologique (BMS), 66, 60–75.

Borgers, N. and Hox, J.P. (2000). Reliability of Responses in Self-Administered Questionnaires: Research with Children. In Social Science Methodology in the New Millennium. Proceedings of the Fifth International Conference on Logic and Methodology, Cologne, Germany, J. Blasius, J.P. Hox, E. DeLeeuw, and P. Schmidt (eds). (CD-Rom).

Borgers, N. and Hox, J.P. (2001). Item Non-Response in Questionnaire Research with Children. Journal of Official Statistics, 17, 321–335.

Borgers, N., Hox, J.P., and Sikkel, D. (2003a). Response Effects in Surveys on Children and Adolescents. The Effect of Offering Labelled Response Options and Vague Quantifiers. In Questioning Childrens' Responses, N. Borgers (ed.). Utrecht: Borgers.

Borgers, N., Hox, J.P., and Sikkel, D. (2003b). Response Quality in Survey Research with Children and Adolescents. The Effect of Response Options, Negative Wording and Neutral Midpoints. In Questioning Childrens' Responses, N. Borgers (ed.). Utrecht: Borgers.

Case, R. (1992). The Mind's Staircase. Exploring the Conceptual Underpinnings of Children's Thoughts and Knowledge. Hillsdale: Erlbaum.

DeLeeuw, E. and Otter, M. (1995). The Reliability of Children's Responses to Questionnaire Items: Question Effects in Children's Questionnaire Data. In Advances in Family Research, J.P. Hox, B.F. van der Meulen, J.M. Kanssens, J.J. ter Laak, and L.W.C. Tavecchio (eds). Amsterdam: Thesis Publisher.

DeLeeuw, E. (1993). Mode Effects in Survey Research: A Comparison of Mail, Telephone, and Face-to-Face Surveys. Bulletin de Methodologie Sociologique, 41, 3–19.

Demetriou, A. (2000). Organization and Development of Self-Understanding and Self-Regulation. In Handbook of Self-Regulation, M. Boekaerts, P.R. Pintrich, and M. Zeidner (eds). New York: Academic Press.

Demetriou, A., Christou, C., Spanoudis, G., and Platsidou, M. (2002). The Development of Mental Processing: Efficiency, Working Memory, and Thinking. Monographs of the Society for Research in Child Development, Serial no. 268, 67, No. 1. Boston: Blackwell Publishing.

Demetriou, A., Efklides, A., and Platsidou, M. (1993). The Architecture and Dynamics of Developing Minds. Experimental Structuralism as a Frame for Underlying Cognitive Developmental Theories. Monographs of the Society for Research in Child Development, Serial no. 234, Vol. 58, No. 5–6. Boston: Blackwell Publishing.

Fuchs, M. (2003). Data Quality in Surveys Among the Youth. Item Non-Response and Response Errors. Paper presented at the International Conference on Methoden und Techniken der Kindheitsforschung, Hohenkammer, April.

Fuchs, M., Lamnek, S. and Luedtke, J. (2001). Tatort Schule. Gewalt an Schulen 1994–1999. Opladen: Leske + Budrich. [In German]

Fuchs, M., Lamnek, S., and Wiederer, R. (2003). Querschläger. Jugendliche zwischen rechter Ideologie und Gewalt. Opladen: Leske + Budrich. [In German]

Goswami, U. (2002). Blackwell Handbook of Childhood Cognitive Development. Malden: Blackwell.

Groves, R.M. (1989). Survey Errors and Survey Costs. New York: Wiley.

Gubrium, J.F. and Holstein, J.A. (2002). Handbook of Interview Research. Context and Method. Thousand Oaks: Sage.

Hershey, M.R. and Hill, D.B. (1976). Positional Response Set in Pre-Adult Socialisation Surveys. Social Science Quarterly, 56, 707–714.

Herzog, A.R. and Rodgers, W.L. (1992). The Use of Survey Methods in Research on Older Americans. In The Epidemiologic Study of the Elderly, R.B. Wallace and R.F. Woolson (eds). New York: Oxford University Press.

Hess, J., Rothgeb, J., Zukerberg, A., Richter, K., Le Menestrel, S., Moore, K. and Terry, E. (1998). Teens Talk: Are Adolescents Willing and Able to Answer Survey Questions? Paper presented during a poster session at the Annual Conference of the American Association for Public Opinion Research, May 14–17, St. Louis, Missouri.

Hippler, H.-J. and Schwarz, N. (1992). The Impact of Administration Mode on Response Effects in Surveys. ZUMA-Arbeitsberichte 92/14. Mannheim, Germany.

Holbrook, A.L., Green, M., and Krosnick, J.A. (2003). Telephone Versus Face-to-Face Interviewing on National Probability Samples with Long Questionnaires. Public Opinion Quarterly, 67, 79–125.

Knäuper, B. (1999). The Impact of Age and Education on Response Order Effects in Attitude Measurement. Public Opinion Quarterly, 63, 347–370.

Knäuper, B., Belli, R.F., Hill, D.H., and Herzog, A.R. (1992). The Use of Survey Methods in Research on the Older Americans. In The Epidemiologic Study of the Elderly, R.B. Wallace and R.F. Woolson (eds). Oxford: Oxford University Press.

Kohlberg, L. (1976). Moral Stages and Moralization. The Cognitive-Developmental Approach. In Moral Development and Behavior. Theory, Research and Social Issues, T. Lickona (ed.). New York: Holt, Rinehart and Winston.

Krosnick, J.A. (1992). The Impact of Cognitive Sophistication and Attitude Importance on Response Order and Question Order Effects. In Context Effects in Social and Psychological Research, N. Schwarz and S. Sudman (eds). New York: Springer.

Krosnick, J.A. and Alwin, D. (1987). An Evaluation of Cognitive Theory of Response-Order Effects in Survey Measurement. Public Opinion Quarterly, 51, 201–219.

Lyberg, L., Biemer, P., Collins, M., DeLeeuw, E., Dippo, C., Schwarz, N., and Trewin, D. (eds) (1997). Survey Measurement and Process Quality. New York: Wiley.

Lyberg, L. and Kasprzyk, D. (1991). Data Collection Methods and Measurement Error: An Overview. In Measurement Error in Surveys, P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman (eds). New York: Wiley.

Marsh, H.W. (1986). Negative Item Bias in Rating Scales for Preadolescent Children: A Cognitive-Developmental Phenomenon. Developmental Psychology, 22, 37–49.

Miller, W. (1976). Preschoolers' Attitudes Toward Libraries: Can They Be Measured? American Libraries, 7, 151–153.

Piaget, J. (1948). The Moral Judgment of the Child. Glencoe, Ill: Free Press.

Schuman, H. and Presser, S. (1981). Questions and Answers in Attitudes Surveys. Experiments on Question Form, Wording, and Context. Thousand Oaks: Sage.

Schwarz, N. and Hippler, H.-J. (1995). The Numeric Values of Rating Scales: A Comparison of Their Impact in Mail Surveys and Telephone Interviews. International Journal of Public Opinion Research, 7, 72–74.

Schwarz, N., Hippler, H.-J., Deutsch, B., and Strack, F. (1985). Response Scales. Effects of Category Range on Reported Behaviour and Comparative Judgements. Public Opinion Quarterly, 49, 388–395.

Schwarz, N., Strack, F., and Mai, H.P. (1991a). Assimilation and Contrast Effects in Part-Whole Question Sequences: A Conversational Logic Analysis. Public Opinion Quarterly, 55, 3–23.

Schwarz, N., Hippler, H.-J., and Noelle-Neumann, E. (1992). A Cognitive Model of Response Order Effects in Survey Measurement. In Context Effects in Social and Psychological Research, N. Schwarz and S. Sudman (eds). New York: Springer.

Schwarz, N. and Knäuper, B. (2000). Cognition, Aging and Self-Reports. In Cognitive Aging. A Student Primer, D. Park and N. Schwarz (eds). Philadelphia: Psychology Press.

Schwarz, N., Knäuper, B., Hippler, H.-J., Noelle-Neumann, E., and Clark, L. (1991b). Rating Scales. Numeric Values May Change the Meaning of Scale Labels. Public Opinion Quarterly, 55, 570–582.

Scott, J. (1997). Children as Respondents: Methods for Improving Data Quality. In Survey Measurement and Process Quality, L. Lyberg, P. Biemer, M. Collins, E. DeLeeuw, C. Dippo, N. Schwarz, and D. Trewin (eds). New York: Wiley.

Sudman, S., Bradburn, N., and Schwarz, N. (1996). Thinking about Answers: The Application of Cognitive Processes to Survey Methodology. San Francisco: Jossey Bass.

Tourangeau, R., Rips, L.J., and Rasinski, K. (2000). The Psychology of Survey Response. New York: Cambridge University Press.

Villancourt, P.M. (1973). Stability of Children's Survey Responses. Public Opinion Quarterly, 37, 373–387.

Zukerberg, A. and Hess, J. (1997). Uncovering Adolescent Perceptions: Experiences Conducting Cognitive Interviews with Adolescents. Paper presented at the Annual Meeting of the American Association of Public Opinion Research, Norfolk, VA.

Received November 2002
Revised November 2004