

Social Science Computer Review

<http://ssc.sagepub.com/>

Differences in the Visual Design Language of Paper-and-Pencil Surveys Versus Web Surveys: A Field Experimental Study on the Length of Response Fields in Open-Ended Frequency Questions

Marek Fuchs

Social Science Computer Review 2009 27: 213 originally published online 24 February 2009

DOI: 10.1177/0894439308325201

The online version of this article can be found at:

<http://ssc.sagepub.com/content/27/2/213>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Social Science Computer Review* can be found at:

Email Alerts: <http://ssc.sagepub.com/cgi/alerts>

Subscriptions: <http://ssc.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://ssc.sagepub.com/content/27/2/213.refs.html>

>> [Version of Record](#) - May 7, 2009

[OnlineFirst Version of Record](#) - Feb 24, 2009

[What is This?](#)

Differences in the Visual Design Language of Paper-and-Pencil Surveys Versus Web Surveys

A Field Experimental Study on the Length of Response Fields in Open-Ended Frequency Questions

Marek Fuchs

University of Kassel, Germany

Respondents in self-administered, paper-based surveys answer open-ended frequency questions differently depending on the size of the response field provided. Longer fields increase the proportion of alphanumeric elaborations or answers explicitly designated as an estimate (“about 20,” “60-70”). By contrast, shorter fields yield higher proportions of pure digit responses. It is assumed that respondents interpret the size of the response field as an instruction regarding the format of the expected answer. In this article, the authors assess the effect of the length of the response field in a web survey compared to a paper-and-pencil questionnaire. Results indicate that the effect of the length of the response field is more pronounced in the paper-and-pencil condition compared to the web survey condition. This raises the question as to what extent the underlying visual design language differs across self-administered survey modes.

Keywords: *web survey; visual design language; frequency question*

Asking frequency questions—like, “On average how many cigarettes do you smoke a day?”¹—in self-administered surveys requires the researcher to decide on a crucial issue: The question could either be designed as close-ended, providing frequency ranges as response options, or, by contrast, be designed as an open-ended frequency question, which offers a response field and expects respondents to indicate digits. Open-ended frequency questions are often preferred because in close-ended frequency questions, the respondent’s answer is affected by scale effects induced by the range of the response options (e.g., Schwarz, 1999). However, experiences from the postfield editing stage of many surveys suggest that respondents do not only provide pure numbers and digit responses to open-ended frequency questions but also alphanumeric elaborations of their answer (e.g., “depends on the day of the week”) or explicit estimates (e.g., “about 20,” “~ 50,” “60-70”), which are of lower quality and cause additional labor during the data cleaning stage of a survey.

Author’s Note: Correspondence concerning this article should be addressed to Marek Fuchs, University of Kassel, Fachbereich Gesellschaftswissenschaften, Kassel, Germany 34109; e-mail: marek.fuchs@uni-kassel.de

Results from earlier studies (Christian, Dillman, & Smyth, 2007; Couper, Traugott, & Lamias, 2001; Couper, Tourangeau, & Conrad, 2007; Smyth, Dillman, & Christian, 2007) indicate that the size and the design of the response field associated with an open-ended question may influence the response behavior. In general, larger response fields evoke more information from the respondents, which in the cases of open-ended frequency questions should lead to more alphanumeric elaborations and explicit estimates.

Even though alphanumeric elaborations provided by the respondent may clarify the response and thus increase data quality, often such responses do not denote a specific number, which in turn increases the uncertainty when coding them in the postfield stage of a survey. Also because the postfield editing of alphanumeric elaborations is labor-intensive, questionnaire designers aim to reduce the occurrence of such respondent behavior. The same is true for explicit estimates (e.g., "approx. 25"), but because they include a digit response, coding is more reliable. Furthermore, explicit estimates provide additional information regarding the underlying response strategy applied by the respondent ("estimation" vs. "recall and count"; Sudman, Bradburn, & Schwarz, 1996), which might be of some value for the researcher when interpreting the results. Despite those possible advantages, most of the time researchers prefer pure digit responses, which are easier to handle and presumably of higher quality. This article aims at the development of general guidelines for questionnaire designers on how to collect pure digit responses with open-ended frequency questions.

Previous research on formal characteristics of response fields has been conducted either completely in a web survey environment (e.g., Christian, Dillman, & Smyth, 2005, 2007; Couper et al., 2001; Fuchs, 2007a) or solely in paper-based, self-administered surveys (e.g., Dillman, Gertseva, & Mahon-Haft, 2005). In this article, we will test whether the same manipulation of the formal characteristics of the response field yields similar effects in both survey modes. In particular, we will focus on the effect of the response field's size in a 2×2 -factorial design, comparing the effect of a small box versus a large box in the paper-and-pencil portion of a survey and in a web-based portion of the same study. Thereby we will contribute to the mode-specific knowledge on the visual design language in self-administered surveys.

Background and Hypothesis

Based on the question-answer process as described by Sudman and colleagues (1996), it is assumed that respondents interact in a survey interview according to everyday rules of communication. In particular, the respondents' contributions to the survey interview are expected to be clear, informative, relevant, and truthful. Based on these rules, respondents react to the questions and other verbal contributions of the interviewer. These general assumptions have motivated a line of methodological research regarding the impact of the wording of a question and of its response categories on the answers obtained from the respondent (e.g., Sudman et al., 1996; Tourangeau, Rips, & Rasinski, 2000). In self-administered surveys, no interviewer contributes to the interaction. Instead, the questionnaire comprises all contributions of the researcher to the interaction with the respondent. Because respondents interact with the questionnaire themselves, they do not only adhere

to the verbal language printed in the questionnaire but also to formal aspect of the design of each question and of the questionnaire in general.

Several recent experiments on the visual design aspects of self-administered surveys demonstrate that graphical elements of a survey question may influence its cognitive processing and thus the answers obtained from respondents (Couper et al., 2007). Evidence from this line of research in the framework of visual design effects (e.g., Christian & Dillman, 2004; Christian et al., 2005, 2007; Couper et al., 2001; Smyth et al., 2006, 2007; Spörrle, Gerber-Braun, & Försterling, 2007) indicates that the length of a response field may have a dramatic impact on the question-answer process and on the responses provided.

In an early (unintended) web experiment, Couper and colleagues (2001) compared a short response field with a long one for a question that required respondents to report numbers between 1 and 10. Even though this experiment was not planned (it rather resulted from a programming error), they found significantly more respondents who entered an invalid response in the long entry box version than in the short entry box version. Furthermore, a detailed inspection of the data revealed that respondents were influenced by the size of the entry box and provided more information than was required. For example, instead of simply providing a number, respondents would enter "about 3" or "between 4 and 5" (Couper et al., 2001).

Later, Christian and Dillman (2004) showed that the space provided with an open-ended text question is perceived as a meaningful signal by the respondent. "In two of the three questions, the larger answer space significantly produced a greater number of themes and topics mentioned in the answer" (Christian & Dillman, 2004, p. 68). Based on their results, respondents appear to interpret the space provided as an indicator of the importance of a question as well as the expected effort when responding to the question (see also Dillman et al., 2005).

Similar results were found for open-ended, list-style questions (Smyth et al., 2007). Again systematic and theoretically predictable effects were found indicating that respondents interpreted the size and structure of the response field (one large box vs. multiple small boxes) as relevant instructions and modified their response behavior accordingly. Spörrle and colleagues (2007) were able to replicate several of these findings. Thus we speculate that the length of an open-ended frequency question should affect the answers provided in a similar way.

Support for this assumption comes from research on open-ended questions on temporal information. Christian and colleagues (2007) found that respondents were more likely to provide a verbal version of a month (e.g., "Aug" instead of "08") when the response box was large enough to accommodate extra characters than when it was designed to hold only two digits or characters. Again the size and shape of the response field seem to be interpreted by the respondent and considered when providing an answer.

In the same study, Christian and colleagues (2005) demonstrated that the effects of the visual design might be even stronger than the influence of manipulations of the questions' wording. In addition to the size of the box, the experiment tested the effects of a verbal language change in the question stem ("When did you begin your studies?" vs. "What month and year did you begin your studies?"). The more specific month and year instruction was expected to increase the percentage of respondents reporting their answer in the desired format (two digits for the month and four for the year). However, the modification

of the verbal language had no visible effect on the proportion of answers in the desired format. In general, this finding indicates the importance of visual design effects in relation to verbal language effects and in particular it demonstrates the potential effect of the length of the response field in a web survey.

Hypothesis 1 + 2

In the present article, this line of research on open-ended text questions and questions on dates will be extended to open-ended questions on frequencies and quantities. Particularly, we will focus on questions regarding the number of units of a given type (students, people, money, etc.). In a prior study, we assessed the impact of—among other formal characteristics—the size of the response field for open-ended frequency questions in paper-and-pencil, self-administered surveys (Fuchs, 2007a). In various field experiments, we confirmed the hypothesis that respondents provided more explicit estimates and alphanumeric elaborations when using long response fields compared to short response fields. This was especially true for questions that required respondents to make use of an estimation strategy (instead of “recall and count”; Belli, Schwarz, Singer, & Talarico, 2000). In line with the interpretation provided by Christian and colleagues (2007), this effect seems to occur because of the respondent’s perception of the space provided; the size of the response field was perceived as an indicator for the expected elaborateness of the response.

Hypothesis 1: In line with previous experiments and evidence reported in the literature, we aim to replicate the already known finding according to which the length of the response field has a significant impact on the response properties in a paper-and-pencil survey. When confronted with a long response field, respondents are more likely to report alphanumeric information and explicit estimates in addition to, or instead of, pure digit responses. By contrast, a smaller response field will be perceived as a signal that the answer should be limited to numbers only, which in turn should increase the likelihood of pure digit responses.

Hypothesis 2: Because a web survey shares many properties of a paper-based self administered survey, we predict that web survey respondents—like respondents in paper-and-pencil surveys—are more likely to report alphanumeric information and explicit estimates in response to a larger response field. By contrast, a smaller response field will increase the likelihood of implicit estimates and pure digit responses.

Hypothesis 3

Compared to a paper-based, self-administered survey, the question-answer process in a web survey differs to a certain degree: The questionnaire is answered in a rather segmented manner and each question is treated separately from previous questions (Fuchs, 2002, 2003). In part this is because of the computer-mediated character of an online questionnaire that is strengthened by the interactive approach employed by most web surveys (one item on each screen). For the research reported in this article, it is important to recognize that responding to a web survey requires a computer-mediated interaction, which calls for the use of keyboard and mouse instead of paper forms and pens. Although a paper form allows all kinds of deviations from standard procedures when reporting answer, most users are

aware that computer interfaces are less flexible and less accepting. Although they can write notes and elaborate their responses on the margins of a paper questionnaire (even if the question explicitly asks for a number), most respondents have experienced difficulty when doing similar things on a web page. For example, when placing a bid on an internet auctioning platform, people have learned that they are supposed to enter digits only. If they would enter alphanumeric elaborations in an input field that explicitly asks for a number (e.g., “maximum of 15 Euro”), an error message would appear. Similarly, when answering frequency questions in a web survey, respondents adapt their response behavior to the specifics of the underlying computer-mediated interaction. This assumption raises the question of to what extent the visual design effects for long versus short response fields in paper-based, self-administered surveys apply to web surveys as well.

Christian and colleagues (2005) demonstrated in a comparison of web versus telephone survey that the effects of verbal languages changes are more important in telephone mode while visual design effects occurred—for obvious reasons—in the web survey only. Even though this mode difference involves self-administration and interviewer administration and thus should not be used as a role model for mode differences of two self-administered modes, the results indicate that various modes might have their distinctive visual design language.

Additional evidence for the confined comparability of the visual design language of web surveys and paper-based, self-administered surveys comes from field experimental work on the design of an open-ended frequency question in web surveys (Fuchs, 2007a). Contrary to expectations, the proportion of alphanumeric responses and explicit estimates was considerably smaller compared to prior studies using paper-based questionnaires. Even though these findings were not based on an experimental comparison, we speculated that respondents modify their response behavior in a web survey compared to a paper-based, self-administered questionnaire.

Hypothesis 3: Given the respondent’s presumed expectation that the web interface is less permissive in reaction to ill-formatted responses, we predict a lower level of supplemental alphanumeric input and of explicit estimates in a web survey compared to a paper-and-pencil questionnaire.

Method

The experiment was embedded in a survey on right-wing attitudes conducted among secondary school students in 2001 (see Fuchs, Lamnek, & Wiederer, 2003, for an overview of the substantive results). In a representative study in Bavaria (a state of Germany), 5,042 students (one class randomly selected from a stratified probability sample of 234 Bavarian secondary schools of all types) aged 12 through 21 answered a self-administered questionnaire consisting of some 50 questions (about 200 items) in a classroom setting. The overall response rate was 80%. Female (48%) and male (52%) students were represented appropriately, the mean age was 14.9, and all school tracks were covered according to the respective distribution in the population (lower track: 29%, intermediated track 15%, higher track 29%, vocational schools 26%).

Given the scope of the underlying study, the findings presented in this article are restricted to high school students. Previous research suggests (Fuchs, 2005, 2007b) that

children and juveniles answer survey questions somewhat differently compared to adults. This can be attributed to their limited cognitive abilities (Borgers et al., 2000), as well as to a specific “test-taking mentality,” especially among young children (Hershey & Hill, 1976). Thus we should be careful in generalizing the findings. In addition, because the study was administered in a classroom setting, further restrictions arise. If the paper-based portion of the survey were administered by mail, the less controlled setting could have downgraded respondents’ care and diligence, which in turn might have diminished the visual design effects. Also the person who administered the survey in the classroom (usually the teacher) might have influenced the participants’ response behavior. However, because we asked the teacher to complete some paper work during the administration of the survey, we assume no major effect of the teacher’s presence on the response behavior of the students in a given class.

The experiment employed a 2×2 -factorial randomized between subjects design: questionnaire version 1 versus questionnaire version 2 \times Paper-and-pencil survey versus web survey. A random subgroup of all classes—three quarters of the total sample (172 schools/classes, $N_{\text{paper}} = 3,884$)—answered a paper-based questionnaire. The remaining quarter of the total sample (62 schools/classes, $N_{\text{web}} = 1,158$) completed an identical web questionnaire in the schools’ computer facility also in a classroom setting. Classes were randomly assigned to one of the survey modes and then to either of two questionnaire versions. Questionnaire version 1 (Figure 1, left) provided short response fields for five frequency questions ($N_{\text{paper} + \text{short}} = 2,020$; $N_{\text{web} + \text{short}} = 559$), whereas in questionnaire version 2 (Figure 1, right) long response fields were associated with the same frequency questions ($N_{\text{paper} + \text{long}} = 1,864$; $N_{\text{web} + \text{long}} = 599$). The long response fields had twice the length of the short response fields. Short and long response fields were approximately the same size in both survey modes. It should be noted that the web questionnaire made use of an HTML input field with shaded lines on all four sides, whereas the paper questionnaire displayed a simple response line.

In the paper-and-pencil portion of the survey, randomization was done prior to the field phase; each class was assigned to one of the questionnaire versions. The participants and the persons administering the survey in class were not aware of the various questionnaire versions. Thus we could avoid teachers’ concerns and discussions about the experimentation among students, which potentially could have affected the response behavior. Because teachers and students could not inspect the questionnaire in advance in the web survey condition, respondents were randomly assigned to either of the two versions on an individual basis. It should be noted that the response rates for both survey modes reached fairly high values (80%), which reduces the risk of a mode-specific nonresponse bias.

The survey was conducted predominantly for substantive reasons thus the methodological work discussed here has been enclosed in the study as an addendum. Because of this constraint, the questions used in the experiment could not be developed from scratch; instead, we had to modify existing questions from the respective questionnaires. Five substantive frequency questions were chosen. In addition to the three questions displayed in Figure 1 (Questions 6, 7, and 8), a question on income or amount of pocket money per month (Question S7) and a question on the population of the municipality where the respondent lives (Question S8) were experimentally varied (see Table 1 for a translation of the questions). Aside from the length of the response field, all characteristics of the questions were kept constant across the experimental conditions (wording, position in the questionnaire).

Figure 1**Experimental Question Versions, Short Response Field (Left) Versus Long Response Field (Right) for the Web Survey (Top) and the Paper-Based Survey (Bottom)**

<p>Wie viele Schüler sind in deiner Klasse?</p> <p>Schüler _____</p> <p>Wie viele davon sind Ausländer?</p> <p>Schüler _____</p> <p>Wie viele ausländische Schüler gibt es an deiner Schule ungefähr?</p> <p>Schüler _____</p>	<p>Wie viele Schüler sind in deiner Klasse?</p> <p>Schüler _____</p> <p>Wie viele davon sind Ausländer?</p> <p>Schüler _____</p> <p>Wie viele ausländische Schüler gibt es an deiner Schule ungefähr?</p> <p>Schüler _____</p>
<p>6. Wie viele Schüler sind in deiner Klasse?</p> <p>_____ Schüler</p> <p>7. Wie viele davon sind Ausländer?</p> <p>_____ Schüler</p> <p>8. Wie viele ausländische Schüler gibt es an deiner Schule ungefähr?</p> <p>_____ Schüler</p>	<p>6. Wie viele Schüler sind in deiner Klasse?</p> <p>_____ Schüler</p> <p>7. Wie viele davon sind Ausländer?</p> <p>_____ Schüler</p> <p>8. Wie viele ausländische Schüler gibt es an deiner Schule ungefähr?</p> <p>_____ Schüler</p>

Note: Question 6 (top), “How many students are in your class?” Question 7 (center), “How many of them are of non-German nationality?” Question 8 (bottom), “About how many students of non-German nationality are attending your school?”

To assess the impact of the experimental conditions, responses were entered into a database as written on the paper questionnaire by the respondent (including all alphanumeric characters and blanks). The web questionnaire was designed in a way so that it would accept any input whatsoever, that is, no error message would appear if a respondent entered alphanumeric information into a number field.² From the responses obtained, we extracted the values intended by the respondent. All responses were also coded according to whether the answer given by the respondent was a pure number, an explicit estimate (“about 20,” “~ 50,” “60-70”), or an implicit estimate (heaping and bunching).³ Because heaping and bunching had to be determined based on the range of values in a given response distribution, we applied flexible criteria for implicit estimates for each question depending on the distribution of the responses (details are provided in the note to Table 1). In addition to the explicit and implicit estimates, it was determined whether a response contained any other alphanumeric content—regardless of its substantive content. Finally, we applied a separate code for missing responses.

Results

Length of Response Field in the Paper-and-Pencil Survey

In the paper-based survey, we find significant results confirming previous findings (Table 1, left side): For four of the five frequency questions, we observe a significant larger

Table 1
Percentage of Explicit and Implicit Estimates, Alphanumeric Responses, and Missing Data for Five Survey Questions on Frequencies and Numbers for Short Versus Long Response Fields, Paper-Based Survey Condition (Left), and Web Survey (Right)

	Paper-Based Survey			Web Survey			Diff Short P&P—Web	Diff Long P&P—Web
	Short	Long	Diff	Short	Long	Diff		
Question 6: “How many students are in your class?”								
Numbers only	98	99	-1*	92	79	13***	6***	20***
Pure digits	88	91	-3**	78	67	11***	10***	24***
Implicit estimates	10	8	2*	14	13	1	-4**	-5**
Explicit estimates	1	1	0	0	1	-1*	1*	0
Alphanumeric response	0	0	0	1	1	0	-1	-1*
Missing data	1	1	0	8	19	-11***	-7***	18***
Question 7: “How many of them are of non-German nationality?”								
Numbers only	84	82	2*	96	96	0	-12***	-14***
Pure digits	77	75	2*	91	87	4*	-14***	-12***
Implicit estimates	7	7	0	5	10	-5**	2*	-3**
Explicit estimates	2	2	0	2	1	1	0	1
Alphanumeric response	13	15	-2*	2	2	0	11***	13***
Missing data	2	1	1*	1	1	0	1	0
Question 8: “About how many students of non-German nationality are attending your school?”								
Numbers only	69	58	11***	89	88	1	-20***	-30***
Pure digits	51	45	6***	68	65	3	-17***	-20***
Implicit estimates	18	12	6***	22	23	-1	-4 +	-11***
Explicit estimates	11	20	-9***	4	4	0	7***	16***
Alphanumeric response	2	3	1**	1	2	-1	1*	2*
Missing data	18	20	-2 +	6	7	-1	12***	13***
Question S7: “What is the population of the municipality where you live?”								
Numbers only	93	88	5***	93	96	-3**	0	-8***
Pure digits	86	81	5***	80	88	-8***	6***	-7***
Implicit estimates	7	6	1	13	9	4*	-6***	-3*
Explicit estimates	6	10	-4***	4	2	2*	2*	8***
Alphanumeric response	0	0	0	0	0	0	0	0
Missing data	1	2	-1**	3	2	1	-2***	0
Question S8: “How much money do you have available each month (pocket money or earnings)?”								
Numbers only	91	85	6***	91	95	-4**	0	-10***
Pure digits	72	68	4**	74	76	-2	-2	-8***
Implicit estimates	18	17	1	17	19	-2	1	-2
Explicit estimates	8	11	-3***	4	2	2	4**	9***
Alphanumeric response	1	2	-1*	1	1	0	0	1**
Missing data	1	2	-1*	4	2	-2	-3***	0

Note: The p values are results of a chi-square test comparing the percentage of the various response properties in a short versus long response field and paper-based versus web survey, respectively, in 2×2 -tables. Effect sizes for significant differences are between .04 and .17. $N_{\text{total}} = 5,042$; $N_{\text{paper} + \text{short}} = 2,020$; $N_{\text{paper} + \text{long}} = 1,864$; $N_{\text{web} + \text{short}} = 559$; $N_{\text{web} + \text{long}} = 599$; Diff = difference; P&P = paper-and-pencil; Missing data = explicit don't know and no answer; alphanumeric response = word, characters, frequencies, and percentages expressed by words; explicit estimates = ranges, “about,” “approx.,” “~”; implicit estimates = numbers with 2, 3, or 4 zeros in the end depending on the concept of the question; pure digits = digits only, but no implicit estimate.
 * $p < .05$. ** $p < .01$. *** $p < .001$.

proportion of pure digit responses in the short field condition. With the exception of Question 6, respondents who were confronted with a long response field provided pure digits to a lesser extent. Furthermore, the proportion of implicit estimates (heaping and bunching) is larger in the short field condition also for four of the five questions (not significant for Questions S7 and S8; no effect for Question 7). By contrast, a short response field yielded less explicit estimates compared to a long response field. This is especially true for Questions S8, S7, and 8. In addition, for Questions 7, 8, and S8, we observe statistically significant fewer alphanumeric elaborations in the short response field condition.

In sum, results for the paper-and-pencil survey indicate that estimated responses are reported explicitly as an estimate if the response field indicates that such response behavior is acceptable and welcome. A longer field is more likely to be perceived in this way. By contrast, if the response field indicates by its length that pure digits are expected, respondents tend to report their answers more often as pure digits and implicit estimates. In the light of these findings, hypothesis 1 is confirmed.

Based on these findings, the visual design seems to affect the reporting stage of the question-answer process because respondents provide their answers in a different format. However, it remains arguable whether the design also changes the underlying retrieval strategy from estimation to recall and count. The relative size of the increases of pure digit responses and implicit estimates provides some preliminary insight. If it was just a modification of the response behavior on the reporting stage of the question-answer process, the decrease in the proportion of explicit estimates and alphanumeric elaborations should predominantly be accompanied by an increase of the implicit estimates (shift from one type of estimate to another). Instead, results (Table 1, third column) indicate that the decrease of the explicit estimates and alphanumeric elaborations in the short field condition is predominantly attended by an increase of the pure digit responses. The increase of the implicit estimates is small or not significant (with the exemption of Question 8, where the increase of the pure digit responses and the implicit estimates is 6 percentage points each). Thus, in our view, the visual design of the short response field in the paper-and-pencil survey seems to motivate respondents to think harder and to retrieve a more detailed pure digit response. Nevertheless, estimated responses cannot be avoided completely in the short field condition; however, they occur less often (the joint proportion of implicit and explicit estimates is smaller in the short field condition for three of the five questions) and respondents are likely to report them as implicit estimates in the short field condition. Given the magnitude of the effects, the shift in the underlying response process from estimation to recall and count seems to be stronger than the shift from explicit to implicit estimates on the reporting stage of the question-answer process.

Response Behavior in the Web Survey and in the Paper-and-Pencil Survey

This experiment provides the opportunity to assess the characteristics of responses in the web survey condition in comparison with properties of the answers to a paper-based self-administered questionnaire (see Table 1, most right columns). Looking at the explicit estimates first, we detect a striking result: For three of the five questions (Questions 8, S7, and S8), the proportion of explicit estimates is significantly smaller in the web survey condition compared to the paper-based survey condition (on average about 8 percentage

points). For these three questions—that should predominantly be prone to estimation strategies (Questions 8, S7, and S8)—the proportion of explicit estimates is 2 to 5 times larger in the paper-based survey condition. However, for Questions 6 and 7—which call for a “recall and count” response strategy—no visible effects occur.

Looking at the proportions of implicit estimates (heaping and bunching), we detect significantly larger values in the web survey condition for four of the five frequency questions (Questions 6, 8, and S7 for short and long fields, Question 7 for long fields only). Also, the proportion of pure digit responses is usually larger in the web survey condition (Questions 7 and 8 for short and long fields, Questions S7 and S8 for long fields only). In addition, it should be noted that the total proportion of digit responses (implicit estimates plus pure numbers) is significantly larger in the web survey condition for four of the five questions—with the exception of Question 6 where opposite differences occur.

These findings support the hypothesis according to which respondents restricted their responses more often to digits in the web survey mode. In addition, the web survey appears to motivate respondents to a lesser extent to qualify their responses explicitly as a result of an estimation process. In our view, respondents assume—based on their prior experience with computer interfaces—that the web survey would not accept any other input other than pure digits. As cooperative respondents who wish to comply with the tasks, they aim to provide answers that fit the desired format. Accordingly, respondents restricted their responses to pure digits and implicit estimates, which confirms hypothesis 3.

Length of Response Field in the Web Survey

To our surprise, we do not find the expected impact of the length of the response field on the proportion of explicit estimates and alphanumeric elaborations in the web survey condition. For both, the long as well as the short response fields, the proportion of explicit estimates and alphanumeric input is nearly identical for all five questions tested in this experiment. There is only one exemption: For Question 6 where we found a significant 1 percentage point difference in the expected direction. However, for all other questions, no significant effect occurs.

For the proportion of implicit estimates, we also do not find the expected results: Even though, the proportion of heaping and bunching is significantly larger in the short field condition for Question S7, for all other questions, we do not see significant differences in the expected direction. Similarly, for the pure digit responses, we observe significant effects in favor of the short field condition for Questions 6 and 7; however, for the other items, no significant difference occurs or it is contrary to expectations (Question S7).

Even when combining pure digit responses and implicit estimates, we do not see the expected effect of the short response field in the web survey condition. With the exemption of Question 6, the long response field condition yields numbers to the same extent (Questions 7 and 8) or even significantly more number of responses (Questions S7 and S8) compared to the short response field condition.

In sum, despite some isolated findings—for example, a considerable increase of item-missing values for the long field condition in Question 6, which we cannot explain—the length of the input field seems to have no major effect on the response properties in the web survey condition. Thus hypothesis 2 could not be confirmed. Presumably, the web survey’s

character as a computer interface overrides the effects of visual design properties such as the response field's length.

A detailed analysis for female and male respondents as well as by age groups confirms the overall findings; even though older respondents seem to answer frequency questions differently (more item missing values, more implicit estimates), the overall effect of the length of the response field is similar. In addition, female and male students differ in the way they respond to frequency questions: female respondents provide more explicit estimates and show more item missing value, whereas male students report more implicit estimates. However, both groups are prone to the same visual design effects in the paper-and-pencil condition and they show similar small or nonexistent effects in the web survey condition.

Discussion

In the paper-based portion of our experiment, we found clear evidence for the predicted impact of the visual design on the properties of the answers provided by the respondents (hypothesis 1). In the long response field condition more alphanumeric elaborations and explicit estimates occurred, whereas the short response field version yielded more pure number responses and implicit estimates. In contrast to the paper-and-pencil condition and unlike other experiments with paper-based surveys (Fuchs 2007a), we found fewer explicit estimates and alphanumeric elaborations in the web survey condition. Accordingly, hypothesis 3 is confirmed.

By contrast, we found little support for hypothesis 2: Surprisingly, in the web survey condition, almost no effect of the length of the response field occurred. In the short field condition, we found an increase of the pure digit responses for two of the five questions and an increase of the implicit estimates for one other question. However, for the majority of the comparisons, no effects in the expected direction occur. It should be noted that the effects discussed in hypothesis 3 leads to very few explicit estimates and alphanumeric elaborations. As a result, the size of the response field had little chance to influence the response behavior in the predicted direction.

Because a web survey requires the respondents to interact with a computer interface rather than with a paper form (which changes basic properties of responding to a question), we speculate that the visual design principles related to the length of the response field may not be fully transferable from paper-based surveys to web surveys. In our view, the character of the web survey as a computer interface prompts respondents to restrict their input to a certain format, which in turn overrides or modifies the effect of the length of the response field.

Interestingly, other studies found notable effects of the length of the response field in web surveys, for example, for temporal questions on dates (Christian et al., 2007). However, although dates can be reported in various suitable and correct formats (e.g., "Aug 08", "Aug 2008", "08/08", "08/2008")—even with computer interfaces—frequencies and quantities are usually reported using digits only (especially when interacting with computers). Thus we assume that the visual design has the opportunity to affect the respondents' decision on how to report dates because multiple conventional options of reporting

such information exist. By contrast, for frequencies and quantities, the spectrum of commonly accepted means of reporting is restricted—especially when interacting with a computer interface.

So far the experiment reported here covers just a small segment of all possible visual design features of web surveys. In previous studies, we have shown that a label associated with the response field decreases the proportion of explicit estimates and alphanumeric elaborations in paper-based surveys. In the same study, we have demonstrated that a response box with placeholders for each digit increases the proportion of pure digit responses. Also, it could be demonstrated that preset values (Fuchs 2007a) and masks for the response fields (see Couper et al., 2007) may enhance the quality of the responses obtained. Given the evidence at hand, it remains to be proven whether web surveys are prone to these effects as well or whether the web survey's dominant character as a computer interface will prevail over the possible impact of those visual design elements.

Our results provide preliminary evidence for questionnaire design rules and supplement existing knowledge on questionnaire design principles: questionnaire designers are supposed to communicate their questions consistently, including the verbal language as well as the visual design of a question. For example, questions asking for a precise number should not provide a lengthy response field. This would be conceived by the respondent as ambiguous information or be treated as an invitation to elaborate on the response.

However, this rule is restricted to a certain type of frequency questions (e.g., the number of people or other units). Other types of questions that ask for numbers, such as questions on the number of events or behaviors in a given reference period, need to be explored in the future. Also, even though the questions used in the experiments vary in their distribution—some questions ask for small numbers, others produce large figures—the data at hand do not allow us to test for the impact of the magnitude of the responses on the observed visual design effects.

The findings reported in this article are prone to additional restrictions that limit generalization: As mentioned in the methods section, the findings presented in this article are restricted to German high school students. Thus to generalize, we need to replicate the experiments with samples from the general public. In addition, the paper-and-pencil condition and the web condition do not mirror each other completely. Although the paper version made use of a response line, the web condition offered a response box (at least, most browsers display the underlying HTML code as a box with a visible frame on all four sides). Whether this may have influenced the experimental comparison is not clear. However, evidence from other experiments using paper-based self-administered surveys demonstrates that—in addition to the size of the response field—its shape affects the properties of the answers obtained. It could be demonstrated that a box with clearly visible placeholders for each digit increases the number of pure digit responses in a paper survey; we can only speculate regarding the effect of a response box without placeholders used in the web survey condition of the experiment reported in this article.

Also the short response field used in this study was longer than the maximum length of the frequencies and numbers expected—both in the paper survey as well as in the web survey condition. It might well be that a short field which provides room for the exact number of digits of the expected response will influence the answers obtained to a greater extent and will yield more pure digit responses in the web survey condition. Thus in future

experiments we aim to provide experimental conditions with response fields of various lengths—from the exact number of digits expected to lengthy fields with plenty of room for additional input.

Future research should also explore the relationship of various verbal language and visual language elements. For example, in our experiment, the label attached to the response field was not very instructive to the respondents. Even though the label mentioned the unit of the desired answer, for example, “students”, it did not encourage the respondents to report the number of students. Thus future research will have to explore the interaction of various components of the visual design (e.g., size and shape of response field) and of elements of the verbal language (e.g., label). Also the relationship of verbal instructions in the question stem and visual design elements needs to be explored in greater detail. Such experiments could contribute to a more integrated view on the optimal design of open-ended frequency questions.

In general, it seems plausible that all instructions to the respondent—regardless of whether they use text or graphical language—need to convey a nonambiguous prompt. However, the interplay of various written or verbal instructions and design elements remains to be discovered. Especially, we need to explore whether a more specific question wording leaves room for the visual design effects of the size, the shape, and of the label of response field to influence the response behavior. Christian and colleagues (2005) reported some preliminary results; however, they are focused on temporal questions on dates. Accordingly, we need to replicate those findings for the type of frequency questions used in this article.

Finally, to further explore our provisional explanation of the low proportion of explicit estimates and alphanumeric elaboration and of the minimal effect of the length of the input field in the web survey condition, we need to boost the web survey’s character as a computer interface. One way of doing this would be to emphasize the strictness and the restrictive character of the web surveys, for example, by prompting the respondents several times with warnings and error messages regarding their responses to the questionnaire. Even though one would not like to preoccupy respondents with such burdensome conditions in a real field setting, it would be reasonable to do so in a methods experiment to investigate the specifics of the visual design language of web surveys.

Not surprisingly, our preliminary results suggest that web surveys and paper-based, self-administered questionnaires differ to a certain degree. Even though they exhibit many similarities with respect to key properties (absence of an interviewer, no audio channel, self-completion), web surveys also differ from paper-based ones in that the former is displayed interactively on a computer screen and data entry is done by respondents in the field using keyboard and mouse. So far, it is not clear what other implications for the visual design language arise from those distinctive features.

Notes

1. Example taken from the WHO Monica questionnaire.

2. In this study, five questions have been experimentally manipulated throughout the questionnaire, three at the beginning (Questions 6, 7, and 8) and two at the end (Questions S7 and S8). Because they were positioned apart and given the small number of experimentally varied questions, we assume that the learning curve can be

neglected. Thus we assume that respondents did not alter their response behavior to a given question because they have experienced no warnings or error messages with prior questions. In addition, the fact that open-ended frequency questions (S7 and S8) printed later in the questionnaire are less prone to alphanumeric responses than the early questions (6, 7, and 8) suggests that a learning curve did not affect the response behavior.

3. Heaping and bunching: Respondents answer a question more often with even numbers 10, 20, 30, 50 than with the uneven numbers in between (Jobe, Pratt, Tourangeau, Baldwin, & Rasinski, 1997; Tourangeau & Smith, 1996).

References

- Belli, F. R., Schwarz, N., Singer, E., & Talarico, J. (2000). Decomposition can harm the accuracy of behavioural frequency reports. *Applied Cognitive Psychology*, *14*, 295-308.
- Borgers, N., de Leeuw, E., & Hox, J. (2000). Children as respondents in survey research: Cognitive development and response quality. *Bulletin de Methodologie Sociologique (BMS)*, *66*, 60-75.
- Christian, L. M., & Dillman, D. A. (2004). The influence of graphical and symbolic language manipulations on responses to self-administered questions. *Public Opinion Quarterly*, *68*(1), 57-80.
- Christian, L. M., Dillman, D. A., & Smyth, J. D. (2005). *Instructing Web and telephone respondents to report date answers in format desired by the surveyor. Technical report 05-067*. Pullman: University of Washington, Social and Economic Sciences Research Center.
- Christian, L. M., Dillman, D. A., & Smyth, J. D. (2007). Helping the respondents get it right the first time: The influence of words, symbols, and graphics in web surveys. *Public Opinion Quarterly*, *71*(1), 113-125.
- Couper, M., Tourangeau, R., & Conrad, F. G. (2007, September 17-19). *How the shape and format of input fields affect answers*. Paper presented at the 2007 Workshop on Internet Survey Methodology, Lillehammer, Norway.
- Couper, M. P., Traugott, M. W., & Lamias, M. J. (2001). Web survey design and administration. *Public Opinion Quarterly*, *65*, 230-253.
- Dillman, D. A., Gertseva, A., & Mahon-Haft, T. (2005). Achieving usability in establishment surveys through the application of visual design principles. *Journal of Official Statistics*, *21*, 183-214.
- Fuchs, M. (2002). Comparing internet and paper and pencil surveys. Question order effects and response order effects as indicators for differences of data collection techniques. *Egyptian Journal of Public Opinion Research*, *3*(1), 21-43.
- Fuchs, M. (2003). Kognitive Prozesse und Antwortverhalten in einer Internet-Befragung [Cognitive processes and response behavior in a web survey]. *Österreichische Zeitschrift für Soziologie*, *28*(4), 19-45.
- Fuchs, M. (2005). Children and adolescents as respondents. Experiments on question order, response order, scale effects and the effect of numeric values associated with response options. *Journal of Official Statistics*, *21*, 701-725.
- Fuchs, M. (2007a, May 16-20). *Asking for numbers and quantities: Visual design effects in web surveys*. Paper presented at the AAPOR Annual Conference, Anaheim, CA.
- Fuchs, M. (2007b, July 29-August 2). Face-to-face interviews with children. Question difficulty and the impact of cognitive resources on response quality. Paper presented at the Joint Statistical Meetings, Salt Lake City, UT.
- Fuchs, M., Lamnek, S., & Wiederer, R. (2003). *Querschläger. Jugendliche zwischen rechter Ideologie und Gewalt* [Ricochets. Juveniles amongst right wing ideology and violence]. Opladen, Cologne, Germany: Leske + Budrich.
- Hershey, M. R., & Hill, D. B. (1976). Positional response set in pre-adult socialization surveys. *Social Science Quarterly*, *56*, 707-714.
- Jobe, J. B., Pratt, W. F., Tourangeau, R., Baldwin, A. K., & Rasinski, K. A. (1997). Effects of interview mode on sensitive questions in a fertility survey. In L. Lyberg, P. Biemer, M. Collins, E. deLeeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey Measurement and Process Quality* (pp. 311-329). New York: Wiley.
- Schwarz, N. (1999). Self-reports. How the questions shape the answers. *American Psychologist*, *54*(2), 93-105.

- Smyth, J. D., Dillman, D. A., & Christian, L. M. (2007, May 16-21). *Improving response quality in list-style open-ended questions in web and telephone surveys*. Paper presented at the Annual Conference of the American Association for Public Opinion Research, Anaheim, CA.
- Smyth, J. D., Dillman, D. A., Christian, L. M., & Stern, M. J. (2006). Effects of using visual design principles to group response options in web surveys. *International Journal of Internet Science*, 1(1), 6-16.
- Spörrle, M., Gerber-Braun, B., & Försterling, F. (2007). The influence of response lines on response behavior in the context of open-question formats. *Swiss Journal of Psychology*, 66(2), 103-107.
- Sudman, S., Bradburn, N., & Schwarz, N. (1996). *Thinking about answers. The application of cognitive processes to survey methodology*, San Francisco: Jossey-Bass.
- Tourangeau, R., Rips, L., & Rasinski, K. (2000). *The psychology of survey response*, Cambridge, UK: Cambridge University Press.
- Tourangeau, R., & Smith, T. (1996). Asking sensitive questions. The impact of data collection mode, question format, and question context. *Public Opinion Quarterly*, 60, 275-304.

Marek Fuchs is a professor for research methods and educational sociology at the department of social sciences at the University of Kassel (Germany). He holds a diploma in sociology from University of Munich (1990), a doctoral degree from University of Kassel (1993), and habilitation from Catholic University of Eichstätt-Ingolstadt (2000). He has done postdoctoral work at the Survey Research Center, University of Michigan (1996/97). His research interests include survey research methods. He may be reached at marek.fuchs@uni-kassel.de.