Designing Input Fields for Non-Narrative Open-Ended Responses in Web Surveys

Mick P. Couper[1], Courtney Kennedy[2], Frederick G. Conrad[3], and Roger Tourangeau[4]

Web surveys often collect information such as frequencies, currency amounts, dates, or other items requiring short structured answers in an open-ended format, typically using text boxes for input. We report on several experiments exploring design features of such input fields. We find little effect of the size of the input field on whether frequency or dollar amount answers are well-formed or not. By contrast, the use of templates to guide formatting significantly improves the well-formedness of responses to questions eliciting currency amounts. For date questions (whether month/year or month/day/year), we find that separate input fields improve the quality of responses over single input fields, while drop boxes further reduce the proportion of ill-formed answers. Drop boxes also reduce completion time when the list of responses is short (e.g., months), but marginally increases completion time when the list is long (e.g., birth dates). These results suggest that non-narrative open questions can be designed to help guide respondents to provide answers in the desired format.

Key words: Web surveys; open-ended responses; input field design; date questions; currency questions.

1. Introduction

In a 1995 paper, Tom Smith demonstrated how seemingly incidental design features affected the answers obtained in paper questionnaires (both self- and interviewer-administered). Many other papers have documented the importance of design details, both for interviewer-administered surveys (e.g., Sanchez 1992; Frazis and Stewart 1998; Hansen et al. 2000) and for self-administered surveys (e.g., Christian and Dillman 2004; Jenkins and Dillman 1997). The development and subsequent proliferation of web surveys has also given rise to a number of studies on design, fueled partly by the new design features available in web surveys, and partly by the ease with which experiments on alternative designs could be conducted. Much of this literature has focused on the design of closed-ended questions with limited response options. Relatively few papers have focused on the design of input fields for open-ended answers, whether for narrative responses (e.g., the biggest problem facing the country) or for more constrained input

[1] University of Michigan and JPSM, P.O. Box 1248, Ann Arbor, MI 48106-1248, U.S.A. Email: mcouper@isr.umich.edu
[2] Abt SRBI, 55 Wheeler St., Cambridge, MA 02138, U.S.A. Email: kennedy@srbi.com
[3] University of Michigan, Institute for Social Research, PO Box 1248, Ann Arbor, MI 48106, U.S.A. Email: fconrad@isr.umich.edu
[4] University of Michigan and JPSM, 1218 LeFrak Hall, College Park, MD 20742, U.S.A. Email: rtourang@survey.umd.edu

(e.g., dates, numbers, currency, names, or places). Among the few exceptions are papers by Christian et al. (2007), Smyth et al. (2009), and Fuchs (2009). Our focus here is on non-narrative open-ended responses.

This article adds to this growing body of literature by focusing on several types of open-ended responses commonly used in web surveys. We begin by clarifying the meaning of the term "open question" and identifying different types of open responses. We then describe a series of experiments that explore different types of open-ended responses and how design can be used to encourage respondents to provide answers in the desired formats.

2.   Background

2.1.   *Open Versus Closed Questions*

The survey research literature (e.g., Fowler 1995; Schuman and Presser 1979, 1981; Sudman and Bradburn 1982; Tourangeau et al. 2000) typically makes a distinction between open and closed questions. But, as Sudman and Bradburn (1982, p. 149) note, "the term 'open question' . . . is a bit misleading, since it is really answers that are left open or closed." They define open questions as "those that are answered in the respondent's own words." Even this leaves a wide range of latitude in what is meant by an open question. While questions that elicit narrative answers (see Fowler 1995) are unambiguously open, there are other types of questions that are less so. Here are two response formats for the same question:

1a) During the past 12 months, how many times have you seen or talked with a doctor about your health?
          _____   times
1b) During the past 12 months, how many times have you seen or talked with a doctor about your health?
          ___None
          ___1
          ___2
          ___3 or more times

Schaeffer and Presser (2003) refer to the first form (1a) as an open frequency question and the second (1b) as a closed frequency question (see also Burton and Blair 1991). Tourangeau et al. (2000, p. 231) also refer to open-ended items in which respondents generate a numerical response (Example 1a).

Depending on the mode of the survey, the two types of open questions may be indistinguishable to the respondent. For example, in interviewer-administered surveys, the response provided by a respondent can be "translated" by the interviewer into a form suitable for coding or keying. In other words, it is the interviewer – not the respondent – who puts the response into its final format. Interviewer probing or clarification can help ensure that the response meets system requirements or analytic objectives. In paper self-administered questionnaires, the processing of responses occurs after the fact. Guidance about the desired format of the answer can be provided to respondents in the form of verbal instructions and visual cues, but the researchers

have little control over the formatting of the response and cannot provide immediate feedback to the respondent in the form of probes or error messages. In computerized self-administered questionnaires – such as computer assisted self-interviewing (CASI), interactive voice response (IVR), or web surveys – the burden of providing a response in the desired format lies with the respondent, with the system imposing varying levels of control. At one extreme, the input can resemble a paper form (e.g., a text box permitting unlimited entry); at the other extreme, respondents can be prevented from proceeding until they enter the desired information in a format acceptable to the system. The degree of control exercised over the actor (whether respondent or interviewer) depends on the technology used and how it is designed and implemented (see Couper 2008).

Our focus here is on one particular mode and technology – self-administered web surveys. In such surveys, the design of the instrument – both what is visible to the respondent (question, response options, input fields, etc.) and what is invisible (edit checks, routing actions, etc.) – has implications for the quality of the responses obtained at the time of survey completion. By quality, we mean whether the response is well-formed – that is, whether the response is recognizable to the system (however that may be defined) and permits the system to take the appropriate action – not whether the response is accurate.

2.2. Types of Open-Ended Responses

Open-ended responses can take several forms, and these have different implications for the design of web surveys. Further, the input fields used to capture these responses can vary. There are two types of input fields available in hypertext markup language (HTML) for web surveys: text boxes and text areas. Text boxes are typically used for more constrained responses, where the numbers of characters displayed and the number of characters accepted by the system are specified by the designer. Text areas allow for longer responses and are typically used for narrative responses. The size of a text area is defined by the number of rows and columns displayed and will expand as needed using scrolling for a maximum of 32,700 characters. These two types of text fields are used for most types of open-ended responses, but there are a number of design options in addition to the type of input field. With this in mind, we can identify several different types of open-ended responses.

2.2.1. Narrative Responses With No Length or Formatting Constraints

The first type are questions that impose no constraints on the length or format of the response. This is the classic open-ended question discussed in the questionnaire design literature, in which respondents are invited to articulate their response using their own words. An example of such a question is "What is the biggest problem facing the country today?" Any coding of the responses is done after the fact, and no validation or checking (other than whether a response was provided or not) is possible at the time of response. Text areas are typically used to capture narrative responses in web surveys.

2.2.2.   Short Verbal Responses With Constraints on Length But Not Format

A second type of open question calls for short text answers. Only the length of the response is constrained. Examples of this type are the industry and occupation questions asked in many surveys. For example, these are from the 2008 American Community Survey:

- "What kind of work was this person doing? (For example: registered nurse, personnel manager, supervisor of order department, secretary, accountant)".
- "What were this person's most important activities or duties? (For example: patient care, directing hiring policies, supervising order clerks, typing and filing, reconciling financial records)".

These questions are often subjected to automated coding after the fact (see, e.g., Speizer and Buckley 1998), and could plausibly be coded in real time for verification with the respondent. The designer needs to decide on the size of the text box, thereby encouraging shorter or longer responses. The verbal cues (such as the examples provided in parentheses) also give the respondent guidance about the type of response desired.

Both these first two types are usually subject to coding after the fact. These response formats promote completeness (provision of sufficient information to permit reliable coding) rather than well-formedness (providing a response in a format that the software can evaluate and act accordingly). The difference between them is in the use of text areas versus text boxes in web surveys.

2.2.3.   Single-word/Phrase Verbal Responses

Many survey questions are looking for one- or two-word answers. Questions of this type differ from narrative questions in imposing more constraints on both the length and format of the answers. In paper surveys, this sort of item would require a brief write-in response, but, in web surveys, they can be done as a lookup (e.g., a drop box or select list in HTML), effectively turning the item into a closed question. Examples of this type include country of origin, prescription medication, medical diagnosis, and so on. Unlike Type 2, an exact match exists in a directory or file, and the program may take an action (e.g., a routing decision or error message) based on the response. The list or database must accommodate alternative forms of response (e.g., "Congo, Democratic Republic of the" vs. "Democratic Republic of the Congo," vs. "DR Congo", etc.), must resolve ambiguities (e.g., is "Congo" the "Republic of the Congo" or the "Democratic Republic of the Congo"?), and must be tolerant of misspellings. The choice of a text box versus a select list for this type of response may depend on how long the list is (e.g., there are over 10,000 prescription medications; see Tourangeau et al. 2004), how dynamic the list is (new medications are constantly being added), how easy it is to type the response versus select from a list (medications likely produce more incorrect spellings than, say, country of origin), and what action (if any) is to be taken on the basis of the response (e.g., if a response will drive a skip, fill, or edit check).

2.2.4.   Frequency or Numeric Responses

A fourth type of open item asks for a frequency or some other numerical answer. The item on doctor visits given earlier is an example of this type. These questions appear

straightforward, but several kinds of responses may not meet the question's requirements. These include ranges (e.g., "5 to 7"), estimates ("About 3"), and other verbal responses ("Never," "A lot," "I don't know," "Three," and so forth). Accepting such responses may make it hard to process the response, while forcing a respondent to enter numbers only may result in missing or inaccurate responses when the respondent's best answer does not fit the required format. Closing up the open response (as in 1b above) is one alternative, but may steer respondents toward a certain range of answers (e.g., Schwarz and Hippler 1987; Sudman et al. 1996). Other questions of this type include feeling thermometer ratings, probability estimates, and other question types requiring entry of a number within a prescribed range.

2.2.5. Formatted Numeric or Verbal Responses

The final type of open-ended question seeks an answer in a conventional format, such as dates. These can be numeric (e.g., telephone number, Social Security number, ZIP code, currency amounts) or character (e.g., 2-character state abbreviation, e-mail address), or a mix of the two (e.g., dates such as January 1, 2010). The conventional format may include delimiters or special characters (many of the above examples can be entered with or without such additions). Thus, even with the existence of a recognizable convention, considerable latitude may exist (e.g., U.S. phone numbers as (xxx)-xxx-xxxx, or xxx-xxx-xxxx, currency amounts as xxx or $x,xxx.xx, etc.)

There are several possible ways to design questions to elicit such responses. For example, a single input box could be provided or a separate box provided for each component (such as month, day, and year of birth). Some (like date of birth) could be designed using drop boxes. In all cases, the format of the response can be assessed on the spot, generating an error message for responses that do not follow the convention. If a particular format of response is desired, the use of templates and verbal instructions may encourage well-formed responses. Templates are verbal labels, special characters (such as slashes or hyphens) or symbols (such as the dollar sign) associated with a text box to clarify the format of the material to be entered.[5]

In summary, open-ended responses vary in the degree and type of constraints they impose. Items that elicit verbal answers (our first three types of open question) can impose no constraints at all (allowing lengthy narrative responses), constrain the respondent to short verbal answers, or constrain the responses to choices that could, in principle, be provided in a look-up table. Items that elicit numerical or formatted responses (our final two types) can impose minimal constraints on the format of the answer (as with items that ask only for whole numbers) or demand that the response follow specific formatting conventions (such as the MM/DD/YYYY convention for dates in the U.S.). In interviewer-administered surveys, interviewers can ensure that the answers meet the formatting requirements, but in self-administered (and particularly computerized) surveys, these requirements must be communicated to the respondent.

---

[5] Templates can be viewed as using both symbolic and numeric language in Redline and Dillman's (2002) terminology (see also Christian and Dillman 2004).

The survey literature makes it clear that respondents may be unaware of the desired format of the open-ended response and are likely to vary in how they format their answers. The goal of good instrument design is thus to reduce ambiguity about what is required, thereby reducing variation in the format of entries, making both respondents' and analysts' tasks easier.

2.3. Prior Research on Open-Ended Responses in Web Surveys

While there are several studies contrasting questions eliciting open-ended responses on the web to similar questions in other modes, such as mail (see, e.g., DeMay et al. 2002; Elig and Waller 2001; MacElroy et al. 2002), the literature on the *design* of such questions in web surveys is relatively sparse. Several studies have examined the effect of the length of the input field on narrative responses. Dennis et al. (2000) found that longer fields encouraged longer responses. Smyth et al. (2009) found that increasing the size of input fields had no effect on early respondents, but did increase response quality (length of responses and number of themes) among late respondents.

The effect of input field size on non-narrative responses is more mixed. Couper et al. (2001) found that respondents answering frequency questions were significantly more likely to provide ranges, qualified responses, and other nonnumeric input when the input field was larger (about 16 spaces) than when it was smaller (2 spaces). Christian et al. (2007) experimentally varied a number of design features in questions eliciting dates (month and year) in a survey among college students. They found that the size of the answer spaces, the style of labels used (words versus symbols), and the alignment of the labels and answer spaces "independently and jointly increase the percentage of respondents reporting their answer in the desired format," where the desired format was a 2-digit month and 4-digit year (Christian et al. 2007, p. 123).

Fuchs has conducted a series of experiments on the design of such questions. In one, he varied the size of input fields for a series of frequency questions in both paper and web surveys administered to high school students (Fuchs 2009). He found significant effects of the length of the input field on responses in the paper version (with longer fields producing more ill-formed answers), but not for the web version. In a second experiment, Fuchs (2007) varied both the length of the input field and the use of labels in a web survey among college students. Again he found no differences between the long and short input fields in whether the desired numeric format was entered. He also found no effect of labels on the type of answers provided. In a third web experiment, Fuchs (2007) experimented with placing a correctly formatted default value ("0,000.00") inside the input field for a question on the amount spent on alcoholic beverages. Those who got the default value were significantly more likely than those who did not get the default to enter pure numbers (34% versus 17%) and correspondingly less likely to enter alphanumeric values (6% versus 22%), suggesting that such defaults guide the respondent to the desired form of input.

In summary, a couple of studies suggest that longer input boxes may reduce the proportion of respondents who give correctly formatted numerical answers (Couper et al. 2001; Christian et al. 2007), but two studies by Fuchs failed to find similar effects for the length of the input box. While the populations were similar (college or high school

students), the studies varied in the types of questions asked and in the experimental manipulations used. The mixed results on the effect of length of input field on the format of frequency responses suggest a need for further research. Both Christian et al. (2007) and Fuchs (2007; 2009) found that other features of the input fields (such as the placement of and design of labels and the use of prefilled answers) affected the well-formedness of responses, and we extend this work here.

In the next section, we describe a set of experiments that explore the effects of the size of the input field further. The items in these experiments involve behavioral frequencies and thus pit the conceptual constraints implicit in the question against the visual cues provided by the input box. We also present results from experiments on formatted numeric or verbal responses such as dates.

3.   Design of Experiments and Methods

The experiments described below were embedded in two different web surveys completed by members of opt-in panels. The first survey (Survey 1) was fielded from August 29th through September 13th, 2007. Respondents were drawn from two different on-line panels. The first was the Survey Sampling International (SSI) Survey Spot Panel, an opt-in web panel of volunteers who have signed up on-line to receive survey invitations. The standard SSI incentive (sweepstakes drawing for cash prizes totaling $10,000) was used. Respondents were invited to participate by email and nonresponders received one reminder. A total of 99,381 panelists were invited to participate in the survey, of which 1,200 completed the survey, for a participation rate of 1.2% (see Callegaro and DiSogra 2008). The second panel was the e-Rewards panel, a similar opt-in panel whose members have agreed to complete online surveys, in exchange for receiving e-Rewards currency that can be used to purchase goods and services. Again, respondents were invited to participate by e-mail and nonresponders received one reminder. A total of 31,918 panelists were invited, with 1,206 completing the survey for a 3.8% participation rate. We combine the data from the 2,406 completed surveys for analysis, as the results for the two samples were very similar. An additional 519 Survey Spot and 445 e-Rewards panelists started the survey but did not complete it. These cases are excluded from the analysis; including them does not alter any of the major conclusions presented below.

The second survey (Survey 2) was conducted from April 29th, 2008 to May 14th, 2008. The sample again came from two sources, the first being SSI's Survey Spot Panel. A total of 1,200 panelists completed the survey (out of 30,179 invited), for a participation rate of 4.0%. An additional 328 started the survey but did not complete it. The other cases were from Authentic Response, which maintains a similar opt-in panel and uses sweepstakes entries as incentives. A total of 17,889 members of this panel were invited, with 1,200 completing the survey for a participation rate of 6.7%. An additional 235 started the survey but did not complete it. Again, the data from the 2,400 respondents from the two sample sources who completed the survey were combined for analysis.

Both surveys were programmed and fielded for us by Market Strategies, Inc. (MSI). The experiments described below were included along with several other methodological experiments in these surveys. All experiments were independently randomized, that is,

the assignment of a respondent to a treatment in one experiment was independent of their assignment in another experiment. We tested for, and found no evidence of, carryover effects across experiments.[6]

These are not probability samples permitting inference to the broader population. Rather, they are large and diverse groups of volunteers, and our focus is on the effects of the experimental manipulations on differences in the format of responses. In terms of demographics, 45% of Survey 1 respondents were male, 79% White, 44% college graduates and 55% were under age 50. For Survey 2, 51% were male, 80% White, 35% college graduates and 41% were under 50 years of age.

Our analyses focus on the extent to which respondents provide "well-formed" responses, by which we mean responses in the desired format (see Christian et al. 2007). We also examined response times as an indicator of the efficiency with which respondents answered questions.[7] We describe the design of each of the four sets of experiments, along with the results, in turn below.

4. Experiment 1: Behavioral Frequencies

4.1. Experiment 1 Design

We experimented with the effect of input field size on the reporting of behavioral frequencies (Type 4 above) in two ways, with four different questions. The first experiment (1a) focused on ill-formed responses in the following two questions:

- "In a typical year, how often do you see or talk to a medical doctor?"
- "In a typical year, how often do you go to the dentist?"

One of two input field versions was randomly assigned to respondents – a text box with visible space for four characters, and a text box with visible space for 20 characters (see Figure 1). In both conditions, respondents could enter up to 30 characters.

For most people, the answers are likely to be relatively low numbers and hence unlikely to be produced via estimation (see Conrad et al. 1998; Menon 1993). We might expect a greater effect of the response format in a question where the answer might be subject to estimation; if respondents generate their answers via estimation, the size of the entry box might suggest a larger anchor as a starting point for the estimation strategy. For this reason, we repeated the same experiment on a different set of items where the likelihood of estimation was higher.

The second experiment (1b) used three items from the National Health Interview Survey:

- "How often do you do VIGOROUS activities for AT LEAST 10 MINUTES that cause HEAVY sweating or LARGE increases in breathing or heart rate?"

---

[6] That is, we tested whether the effects of our experimental manipulations were in turn affected by the assignment to treatment conditions in earlier experiments in each survey. We found no significant interactions.
[7] Finally, we also examined breakoffs on these items, but there were very few breakoffs and the rates did not differ significantly across conditions. Thus, we do not discuss breakoffs further in this article.

Fig. 1.   (a) Example of short entry box condition. (b) Example of long entry box condition

- "How often do you do LIGHT OR MODERATE activities for AT LEAST 10 MINUTES that cause ONLY LIGHT sweating or a SLIGHT to MODERATE increase in breathing or heart rate?"
- "In the PAST YEAR, how often did you drink any type of alcoholic beverage?"

The responses were in two parts. First, respondents entered the frequency in a text box, then indicated the time frame for their answers using a set of radio buttons (1 = Day, 2 = Week, 3 = Month, 4 = Year). As with the doctor/dentist visit questions, the size of the text box was varied, with approximately half getting the four-character version, and half the 20-character version, with up to 30 characters permitted in each. In addition, this experiment was crossed with an explicit invitation to provide an estimate, by adding the sentence "Your best guess is fine" to the end of each question, yielding a 2 by 2 design. The two experiments (1a and 1b) were administered twice, in Survey 1 and in Survey 2.

For both sets of questions we hypothesized that the longer input field would encourage more ill-formed responses, including answers in words (e.g., "twice," "never"), ranges (e.g., "4–5") and qualified responses (e.g., "about 3"). For the second set of questions, we hypothesized that the input field size would interact with the permission to estimate, yielding the highest percentage of ill-formed responses in the condition with the longer input field and invitation to estimate. In this case, both the verbal and visual cues may encourage ill-formed responses.

4.2.   Experiment 1 Results

The results from Survey 1 and Survey 2 are very similar, with no difference in substantive conclusions, so we present only the Survey 2 results here (see Table 1).

Manipulating the length of the input box had relatively minor effects on the behavioral frequency reports we examined. In Experiment 1a, respondents receiving the short text box were more likely to enter a well-formed response for the doctor visit item than those

Table 1. Response Quality for Selected Behavioral Frequency Reports by Length of Input Field (Survey 2)

	Experiment 1a: Doctor visits		Experiment 1b: Vigorous exercise	
	Short text box	Long text box	Short text box	Long text box
Well-formed response				
Integer (e.g., 0,1,2,. . .)	99.3%	98.0%	96.8%	97.9%
Ill-formed response	0.1%	0.7%	0.7%	0.8%
Range (e.g., 1 to 2, 3 or 4)	0.0%	0.2%	0.4%	0.5%
Number spelled out (e.g., twice, ten)	0.1%	0.4%	0.2%	0.3%
Other noninteger (e.g., ^, "been a while")	0.0%	0.1%	0.1%	0.0%
No answer	0.7%	1.3%	2.5%	1.3%
Mean response times	15.6 sec.	15.5 sec.	20.3 sec.	21.3 sec.
(Number of observations)	(1,203)	(1,197)	(1,211)	(1,189)

receiving the longer box ($X^2(1) = 6.98, p = .008$), but the proportion of well-formed responses was high among both groups (see Table 1). The difference in well-formed responses failed to reach significance for the dentist visit item ($X^2(1) = 2.02, p = .16$, not shown). The longer text box led slightly more respondents to report a range, such as "3–4 times," or to verbally express a numeric answer, such as "none" or "twice," but the overall rate of such answers is very low.

Turning to Experiment 1b, we found no effect of the invitation to estimate manipulation on the proportion of well-formed responses. We thus collapse over these groups and focus on the text box size manipulation. Contrary to expectation reports about more frequent behaviors – exercise and alcohol consumption – were just as likely to be entered in the intended format as those about rarer behaviors (doctor and dentist visits). Nearly all respondents (97.3%, 97.8%, and 98.6% respectively) provided a well-formed, integer response. The length of the input field had no noticeable effect on whether a well-formed response was provided ($X^2(1) = 2.88, p = .09$ for vigorous activity; $X^2(1) = .58, p = .45$ for light to moderate activity; $X^2(1) = 1.64, p = .20$ for alcohol consumption). The results for vigorous activity are shown in the right-hand side of Table 1; the results for light to moderate activity and alcohol consumption are essentially the same. There was also no relationship between the unit in which the respondent reported an answer (per day, week, month, or year) and the likelihood that the answer was formatted correctly. We also found no effect on the substantive distributions of responses (i.e., median frequencies) by the length manipulation.

The lack of an effect for the length of the input box on these behavioral frequency questions runs counter to the Couper et al. (2001) finding but is consistent with Fuchs's (2007; 2009) web results. The current findings, along with those of Fuchs, suggest that the size of the text box has little effect on responses to questions asking for numeric frequencies, especially when the questions are relatively unambiguous and the answers somewhat straightforward. The implied requirements of the question are clear enough that few respondents provide ill-formed answers, regardless of the potentially misleading cue

provided by the size of the input field. The task in the Couper et al. study was a complex one that asked respondents to distribute ten friends, classmates, etc., across a variety of racial groups.

We also hypothesized that telling respondents that their "best guess is fine" would influence how they formatted their answers. Here again, however, we found that nearly everyone provided a well-formed response. Even when we invited them to estimate and gave them space to enter a range, few respondents entered anything other than a single number.

Neither the length of the input box nor the invitation to provide an estimate had a significant effect on response times for any of the five items examined.

5.   Experiment 2: Currency Amounts

5.1.   *Experiment 2 Design*

The second experiment varied both the size of the input field and the use of templates to elicit a formatted numeric response (Type 5 in our typology). Again, we tested the effects of these variables with two questions:

- "How much did you spend last month on PRESCRIPTION drugs?"
- "How much did you spend last month on NON-PRESCRIPTION or over-the-counter drugs or medications?"

As with Experiment 1, we varied the size of the text box (six characters versus 20 characters). We also varied whether templates were used to encourage well-formed responses. The template consisted of a "$" immediately to the left of the text box and ".00" immediately to the right. This yielded a 2 × 2 design. Two of the four versions are shown in Figure 2. Experiment 2 was included in both surveys.

As with the Experiment 1 items, we defined a well-formed response as the entry of a whole number – that is, no words, ranges, or qualified responses, but also no use of dollar signs, commas, or decimal points. We hypothesized that the shorter text box would

(a)
How much did you spend last month on PRESCRIPTION drugs?

$ [   ] .00

[Next]

(b)
How much did you spend last month on PRESCRIPTION drugs?
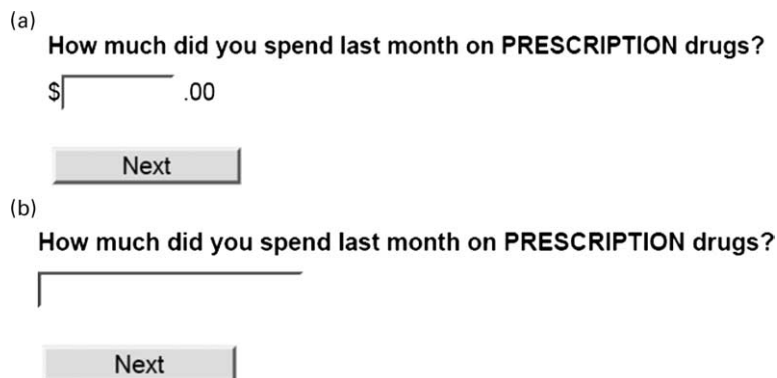
[   ]

[Next]

Fig. 2.   *(a) Short entry box with template. (b) Long entry box without template*

encourage well-formed responses, and that the use of templates would also reduce ill-formed responses, especially those involving dollar signs, commas, and decimals.

5.2. Experiment 2 Results

We found support for both of the above expectations. The results for the prescription and nonprescription drug items from Survey 2 are presented in Table 2; the results for both items from Survey 1 are quite similar. Templates had a stronger effect on how respondents formatted their responses than the length of the text box. When templates were used, 96.0% of responses to the prescription drug item and 96.8% of responses to the nonprescription drug item were entered in the intended format, compared to 81.5% and 80.1% respectively when templates were not used. The main effect of templates is statistically significant for both items ($X^2(2) = 135.8$, $p < .0001$ for prescription drugs; $X^2(2) = 178.5$, $p < .0001$ for nonprescription drugs). In Survey 1, the effect of templates was in the same direction but even larger, with 96% of answers being well-formed with templates versus 73% without. Specifically, templates reduced the use of decimal points and dollar signs in both items.

The influence of the length of the input field was less strong than that of the template, but was still statistically significant ($X^2(2) = 30.6$, $p < .0001$ for prescription drugs;

Table 2. Response Quality for Prescription and Non-Prescription Drug Expenditures by Response Format (Survey 2)

	No template			Template		
	Short box	Long box	Total	Short box	Long box	Total
Prescription drugs						
Well-formed						
Integer (e.g., 50)	86.5%	76.3%	81.5%	98.3%	93.7%	96.0%
Ill-formed	13.0%	23.3%	18.0%	1.4%	5.3%	3.3%
Dollar sign (e.g., $50)	1.5%	3.6%	2.5%	0.0%	0.0%	0.0%
Dollar sign & decimal (e.g., $50.25)	0.6%	1.7%	1.2%	0.0%	0.0%	0.0%
Decimal (e.g., 50.25)	10.6%	17.9%	14.1%	1.2%	5.3%	3.3%
Other (e.g., "too much")	0.3%	0.2%	0.2%	0.2%	0.0%	0.1%
No answer	0.5%	0.3%	0.4%	0.3%	1.0%	0.7%
Mean response time	14.2 sec.	14.2 sec.	14.3 sec.	13.8 sec.	15.0 sec.	14.4 sec.
Nonprescription drugs						
Well-formed						
Integer (e.g., 50)	84.6%	75.5%	80.1%	98.5%	95.2%	96.8%
Ill-formed	14.8%	24.0%	19.3%	1.0%	3.6%	2.3%
Dollar sign (e.g., $50)	1.6%	2.7%	2.2%	0.0%	0.0%	0.0%
Dollar sign & decimal (e.g., $50.25)	0.5%	1.5%	1.0%	0.0%	0.0%	0.0%
Decimal (e.g., 50.25)	12.3%	19.4%	15.8%	0.8%	3.6%	2.3%
Other (e.g., "too much")	0.3%	0.3%	0.3%	0.2%	0.0%	0.1%
No answer	0.6%	0.5%	0.6%	0.5%	1.2%	0.8%
Mean response time	11.5 sec.	11.4 sec.	11.4 sec.	11.5 sec.	11.1 sec.	11.3 sec.
(Number of observations)	(616)	(591)	(1,203)	(587)	(606)	(1,197)

$X^2(2) = 20.4$, $p < .0001$ for nonprescription drugs). In both cases, there were fewer well-formed responses with the longer box (85.2% for prescription and 85.5% for nonprescription drugs) than with the shorter box (92.3% and 91.4% respectively). The effect of length was similar for Survey 1, but not as strong ($X^2(2) = 9.7$, $p = .0077$ for prescription drugs; $X^2(2) = 1.44$, $p = .49$ for nonprescription drugs). This significant effect (for 3 of the 4 items in the two surveys) of the length of the text box runs counter to the results in Experiment 1. The most likely explanation is that Experiment 1 deals with simple frequencies while Experiment 2 deals with currency information, where there is more ambiguity in how the answer should be formatted. In the latter case, respondents differ as to whether they elect to include dollar signs and decimals in their answers. This is clear from the "No Template" condition in Table 2. On average across the two items, some 19% of respondents presented with the long text box and no template decided to include a decimal in their answer, while about 11% of those presented with the short box and no template included a decimal. Logistic regression models were used to test the interaction of templates and input field length; the interaction did not reach statistical significance for either item.

Neither templates nor box length had a significant effect on how quickly respondents completed the items (F(3, 2375) = 0.99, $p = .40$ for prescription drugs; F(3, 2379) = 0.34, $p = .79$ for nonprescription drugs). Finally, we compared the distribution of substantive responses across the conditions, and found no differences.

6. Experiment 3: Month and Year

6.1. Experiment 3 Design

Experiment 3 focuses on formatted numeric responses (Type 5), specifically date information. The experiment was conducted on the following two questions:

- "In what month and year did you last see a medical doctor?"
- "In what month and year did you last see a dentist?"

We tested three different types of input field: a) a single long text box (30 characters), b) two separate fields for the month and year, and c) separate drop boxes for the month and year (see Figure 3). The month drop box presented the months in order from January to December while the year drop box presented them in order from 2007 to 1997, with "Prior to 1997" being the last option. This experiment was implemented in Survey 1.

We repeated this experiment in Survey 2, but replaced the single text box condition with a condition that featured two text boxes, labeled "MM" and "YYYY," allowing us to compare the effects of verbal (Month and Year) and symbolic (MM and YYYY) labels. These questions resemble those used by Christian et al. (2007). Their experiment focused on the two-field version of a question about dates, varying the size of the fields and the labels (words versus symbols), as we did in Experiment 2.

Except for date of birth, questions asking for dates may not be common forms of questions, and strong conventions for responding to them may not exist[8]. Given this, we

---

[8] The American Community Survey and NSF's National Survey of (Recent) College Graduates both have questions of this type (MM/YYYY).

Fig. 3.   (a) Single long input field. (b) Separate fields for month and year. (c) Drop down boxes

would expect the format and type of response options to have a bigger effect on the answers provided than in the earlier experiments.

We expected that separate text boxes would yield more well-formed answers (i.e., a 2-digit month and 4-digit year) than a single text box. We further expected that while the drop boxes would, by definition, yield no ill-formed responses, this might come at the cost of longer completion times and higher rates of missing data.

6.2.   *Experiment 3 Results*

The results for both the doctor visit and dentist visit items from Survey 1 are presented in Table 3. The two-box display and the drop down boxes performed much better than the single text box ($X^2(6) = 374.7, p < .0001$ for last doctor visit; $X^2(6) = 341.5, p < .0001$ for last dentist visit). These response formats featured separate fields for the month and the year, which prompted respondents to report both pieces of information. In the single text box condition, by contrast, nearly one quarter of respondents failed to report either the month or the year of their last doctor visit (22.1%) or dentist visit (22.0%), with year being omitted most often. There is a clear benefit to devoting a separate response field to each piece of information requested from the respondent.

The drop down box format has the added benefit of being faster for respondents to complete. Respondents receiving the drop down boxes responded faster to the doctor visit item than those receiving the text boxes (F(2, 2398) = 13.36, $p < .0001$), although these differences are smaller for the dentist visit item (F(2, 2397) = 3.07, $p = .047$).

In the follow-up study (Survey 2), we replaced the single text box condition with two text boxes, labeled "MM" and "YYYY." Our aim was to determine whether the use of these symbols elicited better-formed responses than the use of the verbal labels "Month" and "Year." This experiment also retained the drop down condition. Table 4 presents the results for the two items.

Table 3. *Response Quality for Month and Year of Last Doctor Visit and Last Dentist Visit by Response Format (Survey 1)*

	One text box	Two text boxes	Drop down boxes
Last doctor visit			
Well-formed response	76.3%	97.7%	99.4%
Verbal month	55.4%	56.4%	0.0%
(e.g., January 2008, Feb 06)			
Numeric month	21.0%	41.3%	99.4%
(e.g., 1/2008, 2 06)			
Incomplete (e.g., month only)	22.1%	0.5%	0.4%
Ill-formed response	1.0%	0.6%	0.0%
No answer	0.5%	1.0%	0.2%
Mean response time	22.7 sec.	24.3 sec.	18.2 sec.
Last dentist visit			
Well-formed response	73.0%	92.4%	96.8%
Verbal month	53.7%	52.3%	0.0%
(e.g., January 2008, Feb 06)			
Numeric month	19.3%	40.0%	96.8%
(e.g., 1/2008, 2 06)			
Incomplete (e.g., month only)	21.7%	2.1%	1.0%
Ill-formed response	4.2%	2.0%	0.0%
No answer	1.0%	3.6%	2.2%
Mean response time	17.3 sec.	18.6 sec.	16.0 sec.
(Number of observations)	(782)	(812)	(812)

Table 4. *Response Quality for Month and Year of Last Doctor Visit and Last Dentist Visit by Response Format (Survey 2)*

	"Month" and "Year" labels	"MM" and "YYYY" labels	Drop down boxes
Last doctor visit			
Well-formed response	96.8%	97.6%	99.4%
Verbal month	54.5%	15.5%	0.0%
(e.g., January 2008, Feb 06)			
Numeric month (e.g., 1/2008, 2 06)	42.2%	82.1%	99.4%
Incomplete	1.3%	1.1%	0.5%
Ill-formed response	0.81%	1.0%	0.0%
No answer	1.2%	0.4%	0.1%
Mean response time	22.4 sec.	23.8 sec.	17.1 sec.
Last dentist visit			
Well-formed response	92.1%	92.8%	95.0%
Verbal month	51.2%	12.3%	0.0%
(e.g., January 2008, Feb 06)			
Numeric month (e.g., 1/2008, 2 06)	40.9%	80.4%	95.0%
Incomplete	3.0%	1.8%	2.1%
Ill-formed response	0.8%	0.8%	0.0%
No answer	4.2%	4.6%	2.9%
Mean response time	17.4 sec.	16.7 sec.	13.2 sec.
(Number of observations)	(770)	(828)	(802)

All three conditions yielded high levels of well-formed responses for both items. If we consider both verbal (e.g., January or Jan) and numeric (e.g., 1 or 01) responses to be well-formed, then the two-box conditions with labels performed equally well. The labels did, however, significantly influence how the month responses were formatted. When the verbal labels were used, the majority of respondents (54.5% for doctor and 51.2% for dentist) spelled out the name of the month. By contrast, when the symbols were used, most respondents (82.1% for doctor and 80.4% for dentist) provided a numeric month. This is consistent with the findings of Christian et al. (2007).

The proportion of well-formed responses did not differ significantly for the verbal and numeric text box label conditions, but there were significantly more well-formed responses in the drop down box condition for both questions ($X^2(2) = 13.94$, $p < .001$ for doctor and $X^2(2) = 5.99$, $p = .050$ for dentist). As in the first study, the mean response times for both doctor and dentist visit items were significantly lower for the drop down condition (F(2, 2375) = 64.23, $p < .0001$ for doctor and F(2, 2377) = 42.98, $p < .0001$ for dentist). For these types of items with a known and limited set of possible response options (12 for month and 15 for year), the drop down box format appears to be superior.

7.   Experiment 4: Date of Birth

7.1.   *Experiment 4 Design*

Questions asking for the respondent's date of birth usually require a formatted numeric or verbal response (Type 5). There are many different ways that researchers can administer a date of birth (DOB) question; for example, they might require only numbers or permit both words and numbers; they might allow delimiters or not allow them; or they might use drop boxes. Also, in the international context, the information on month, day and year of birth can be entered in different orders. However, in the U.S., there is a strong convention – MM/DD/YYYY.[9] Our assumption is if the input format matches this convention, both speed and well-formedness of the response will be facilitated. However, we did not test versions that violated this convention.

We tested four experimental conditions for the DOB question: a) a single long text box (40 characters), b) a single short text box (10 characters), c) three separate text boxes (three characters each for month and day and six for year), and d) three drop boxes. These are illustrated in Figure 4. For the drop boxes, all three lists were presented in ascending order, with years ranging from 1900 to 2000. Again, this experiment was implemented in both surveys.

Our expectation was that the version with three separate text boxes most closely matches the convention and would as a result produce the highest proportion of well-formed responses (MM/DD/YYYY) among the text box versions. Specifically, we expected this version to reduce the use of delimiters and the use of words for month. We also expected the drop box version to take longer to answer than the text box versions, as it requires more clicking and scrolling than the other versions, although by definition

---

[9] Although we note that the forms used by U.S. Customs and Border Protection use a DD/MM/YYYY format.

Fig. 4.   *Date of birth format conditions (clockwise from top left): short text box, long text box, drop down boxes, and three input fields*

the answers would be well-formed. Finally, we expected the long text box version to yield more ill-formed responses (specifically the use of words rather than numbers for month) than the short text box version.

7.2.   Experiment 4 Results

The results for Survey 1 and Survey 2 are very similar, so we present only the latter here (see Table 5). Contrary to expectation, a significantly higher proportion of respondents provided a well-formed birth date in the long text box condition than in the short text box condition ($X^2(1) = 15.90, p < .0001$). In Survey 1, the result was in the same direction though not statistically significant ($X^2(1) = 2.68, p = .10$). The effect of the text box length observed for other questions presumably does not occur with DOB because of the unique properties of this item. For most people, there is little uncertainty about the answer and DOB is a piece of information that one reports fairly often on various forms. The format in which people are requested to report DOB is likely not constant across forms, but the degree of variation is relatively small (e.g., MM-DD-YYYY versus MM/DD/YY). This common experience of reporting DOB in much the same way over and over, leads to consistency in survey reports of DOB, regardless of modest variations in the response format. However, providing three separate input fields significantly improves the proportion of well-formed responses over the single input field versions ($X^2(1) = 38.8, p < .0001$).

Table 5.   *Response Quality for Date of Birth by Response Format (Survey 2)*

	Short text box	Long text box	Three text boxes	Drop down boxes
Well-formed (MM/DD/YYYY)	83.4%	91.1%	96.4%	97.6%
Ill-formed	7.0%	7.0%	1.3%	0.0%
Partially answered	9.2%	1.1%	0.3%	1.0%
Not answered	0.3%	0.8%	2.0%	1.4%
Mean response time	14.6 sec.	15.1 sec.	13.8 sec.	15.9 sec.
(Number of observations)	(585)	(616)	(616)	(583)

As expected, the drop down boxes yielded the highest proportion of responses in the desired MM/DD/YYYY format and no ill-formed responses. One potential drawback is that the drop down boxes for this item took significantly longer to complete than the text boxes ($F(3, 2391) = 7.62$, $p < .0001$). This may be in part due to the larger number of response options in each list relative to the month and year options in Experiment 3, necessitating more scrolling. However, the difference is small, being only 0.8 seconds slower on average than the long text box version. The added effort required to respond did not, however, appear to influence respondents' willingness to answer. The missing data rate with the drop down boxes was similar to that observed with the text boxes.

8. Summary and Discussion

We have examined several different types and design features of questions eliciting nonnarrative open responses in this article. We summarize the results briefly here. First, in contrast to earlier research by Couper et al. (2001) but consistent with that of Fuchs (2007; 2009), we found little effect of the length of the text box on the proportion of respondents giving well-formed responses in questions asking for numeric information, whether frequencies or currency amounts. When the item itself imposes relatively clear requirements on the format of the answer, the length of the text box does not seem to matter. An alternative explanation is that providing an ill-formed answer (such as a range or qualified answer) requires more effort and the opt-in panel members in our studies may have been disinclined to expend such effort, even when encouraged to do so. If the goal is to complete the survey in as short a time as possible, providing a straightforward response is probably the fastest way to do so.

Second, although respondents did not seem to rely on the length of the entry box as a cue regarding the format of their answers, they did seem to make use of the information in the templates in the questions eliciting currency information. Preceding the text box with a "$" and following it with ".00" provides clear cues to the respondent as to what to enter. The templates seem to resolve the ambiguity about whether to include a dollar sign and/or decimal point. When they did not get a template, significantly more respondents provided answers with dollar signs and decimals than when the template was provided. Note that such responses are not incorrect – they simply require more programming effort to convert to a format suitable for analysis.

Third, when a question seeks discrete pieces of information (such as month and year), separate entry boxes yield more well-formed data than an entry field with a single text box. In the latter case, even with instructions provided, the respondent has to figure out how to format the answer; providing separate boxes for each field reduces ambiguity about the desired format. Appropriately labeling such boxes (e.g., with "MM" instead of "Month") further reduces ambiguity about how to answer, as Christian et al. (2007) found. Using a pair of drop boxes where the choice of response options is constrained and the list is relatively short maximizes well-formed answers in less time than text boxes, without increasing item nonresponse. While there are concerns about the use of drop boxes for other types of questions (see Couper et al. 2004), they seem to be optimal for questions of this type, where the order of the items in the lists facilitates finding the appropriate answer rather than influencing what answer to give.

Finally, we find that drop boxes also appear optimal for questions eliciting date of birth. There is a slight time penalty in using the drop boxes for this type of question, in part because the number of choices in each drop box is large. Despite this, item nonresponse rates and ill-formed responses are comparable to those in the condition with three separate text boxes, suggesting that both are acceptable formats for collecting date of birth information. Proving a single text box seems suboptimal for this type of question.

In general, these experiments suggest that, just as respondents have trouble using response scales with closed items (Tourangeau et al. 2004; 2007), they may have trouble deciding what is expected from them with open-ended items. For example, with items calling for a narrative response, respondents may be unsure about the expected length of their answers. With such open-ended items, they seem to infer something about the expected length from the size of the input box (Dennis et al. 2000; Smyth et al. 2009). With other types of open-ended items, they seem to rely on a hierarchy of cues about the intended format of the response, drawing on labels or examples, the number of text boxes, and cues contained in the templates, and only secondarily on the length of the input box. With labels, the format of the label – such as the use of "Month" versus "MM" – may also provide information about the intended format of the answer. Tourangeau et al. (2007) make a similar argument about the cues respondents rely on in using scales; with scales, verbal labels seem to take precedence over numerical labels and numerical labels take precedence over colors. With open-ended items requesting formatted numerical answers, such as dates or currencies, cues like the number of entry fields and the information in the templates help resolve any ambiguities about the desired formats.

Based on these results, we can offer several guidelines for nonnarrative open-ended items. First, if there is a strong convention for a given type of answer, ask for that format. Second, clearly lable the input fields and provide templates to clarify the intended format. Third, if the answer consists of discrete elements, provide a text box for each component or use drop boxes (select lists). And, fourth, provide an appropriate amount of space for reporting an answer. As noted at the outset, these small changes to the format of input fields can affect the well-formedness of the responses provided, reducing error messages (if a well-formed response is required to proceed) and reducing post-survey processing of answers.

9.   References

Burton, S. and Blair, E. (1991). Task Conditions, Response Formulations Processes, and Response Accuracy for Behavioral Frequency Questions in Surveys. Public Opinion Quarterly, 55, 50–79.

Callegaro, M. and DiSogra, C. (2008). Computing Response Metrics for Online Panels. Public Opinion Quarterly, 72, 1008–1032.

Christian, L.M. and Dillman, D.A. (2004). The Influence of Graphical and Symbolic Language Manipulations on Responses to Self-Administered Questions. Public Opinion Quarterly, 68, 57–80.

Christian, L.M., Dillman, D.A., and Smyth, J.D. (2007). Helping Respondents Get it Right the First Time: The Influence of Words, Symbols, and Graphics in Web Surveys. Public Opinion Quarterly, 71, 113–125.

Conrad, F.G., Brown, N.R., and Cashman, E.R. (1998). Strategies for Estimating Behavioral Frequency in Survey Interviews. Memory, 6, 339–366.

Couper, M.P. (2008). Technology and the Survey Interview/Questionnaire. In Envisioning the Survey Interview of the Future, M.F. Schober and F.G. Conrad (eds). New York: Wiley, 58–76.

Couper, M.P., Tourangeau, R., Conrad, F.G., and Crawford, S. (2004). What They See Is What We Get: Response Options for Web Surveys. Social Science Computer Review, 22, 111–127.

Couper, M.P., Traugott, M.W., and Lamias, M. (2001). Web Survey Design and Administration. Public Opinion Quarterly, 65, 230–253.

DeMay, C.C., Kurlander, J.L., Lundby, K.M., and Fenlason, K.J. (2002). Web Survey Comments: Does Length Impact "Quality"? Paper presented at the International Conference on Questionnaire Development, Evaluation and Testing Method, Charleston, SC, November.

Dennis, M., deRouvray, C., and Couper, M.P. (2000). Questionnaire Design for Probability-Based Web Surveys. Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Portland, OR, May.

Elig, T. and Waller, V. (2001). Internet versus Paper Survey Administration: Impact on Qualitative Responses. Arlington, VA: Defense Manpower Data Center, unpublished paper.

Fowler, F.J. (1995). Improving Survey Questions; Design and Evaluation. Thousand Oaks, CA: Sage.

Frazis, H. and Stewart, J. (1998). Keying Errors Caused by Unusual Keypunch Codes: Evidence from a Current Population Survey Test. Proceedings of the American Statistical Association, Survey Research Methods Section, 131–133.

Fuchs, M. (2007). Asking for Numbers and Quantities: Visual Design Effects in Web Surveys and Paper & Pencil Surveys. Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Anaheim, CA, May.

Fuchs, M. (2009). 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. Social Science Computer Review, 27, 213–227.

Hansen, S.E., Beatty, P., and Couper, M.P. (2000). The Effects of CAI Screen Design on User Performance. Paper presented at the Fifth International Conference on Logic and Methodology, Cologne, October.

Jenkins, C.R. and Dillman, D.A. (1997). Towards a Theory of Self-Administered Questionnaire Design. In Survey Measurement and Process Quality, L. Lyberg, P. Biemer, M Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin (eds). New York: Wiley, 165–196.

MacElroy, B., Mikucki, J., and McDowell, P. (2002). A Comparison of Quality in Open-Ended Responses and Response Rates Between Web-Based and Paper and Pencil Survey Modes. Journal of Online Research, 1, <http://ijor.mypublicsquare.com/> .

Menon, G. (1993). The Effects of Accessibility of Information on Judgments of Behavioral Frequencies. Journal of Consumer Research, 20, 431–460.

Redline, C.D. and Dillman, D.A. (2002). The Influence of Alternative Visual Designs on Respondents' Performance with Branching Instructions in Self-Administered Questionnaires. In Survey Nonresponse, R.M. Groves, D.A. Dillman, J.A. Eltinge, and R.J.A. Little (eds). New York: Wiley, 179–193.

Sanchez, M.E. (1992). Effect of Questionnaire Design on the Quality of Survey Data. Public Opinion Quarterly, 56, 206–217.

Schaeffer, N.C. and Presser, S. (2003). The Science of Asking Questions. Annual Review of Sociology, 29, 65–88.

Schuman, H. and Presser, S. (1979). The Open and Closed Question. American Sociological Review, 44, 692–712.

Schuman, H. and Presser, S. (1981). Questions and Answers in Attitude Surveys. New York: Academic Press.

Schwarz, N. and Hippler, H.-J. (1987). What Response Scales May Tell Your Respondents: Information Functions of Response Alternatives. In Social Information Processing and Survey Methodology, H.-J. Hippler, N. Schwarz, and S. Sudman (eds). New York: Springer-Verlag, 163–178.

Smith, T.W. (1995). Little Things Matter; A Sampler of How Differences in Questionnaire Format Can Affect Survey Responses. Proceedings of the American Statistical Association, Survey Research Methods Section, 1046–1051.

Smyth, J.D., Dillman, D.A., Christian, L.M., and McBride, M. (2009). Open-Ended Questions in Web Surveys; Can Increasing the Size of Answer Boxes and Providing Extra Verbal Instructions Improve Response Quality? Public Opinion Quarterly, 73, 325–337.

Speizer, H. and Buckley, P. (1998). Automated Coding of Survey Data. In Computer Assisted Survey Information Collection, M.P. Couper, R.P. Baker, J. Bethlehem, C.Z.F. Clark, J. Martin, W.L. Nicholls, II,, and J. O'Reilly (eds). New York: Wiley, 223–243.

Sudman, S. and Bradburn, N.M. (1982). Asking Questions: A Practical Guide to Questionnaire Design. San Francisco: Jossey-Bass.

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

Tourangeau, R., Couper, M.P., and Conrad, F.G. (2004). Spacing, Position, and Order: Interpretive Heuristics for Visual Features of Survey Questions. Public Opinion Quarterly, 68, 368–393.

Tourangeau, R., Couper, M.P., and Conrad, F. (2007). Color, Labels, and Interpretive Heuristics for Response Scales. Public Opinion Quarterly, 71, 91–112.

Tourangeau, R., Couper, M.P., Galesic, M., and Givens, J. (2004). A Comparison of Two Web-Based Surveys: Static vs Dynamic Versions of the NAMCS Questionnaire. Paper presented at the RC33 International Conference on Social Science Methodology, Amsterdam, August.

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