

Cognitive probes in web surveys: on the effect of different text box size and probing exposure on response quality

Behr, Dorothee; Bandilla, Wolfgang; Kaczmirek, Lars; Braun, Michael

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Dieser Beitrag ist mit Zustimmung des Rechteinhabers aufgrund einer (DFG geförderten) Allianz- bzw. Nationallizenz frei zugänglich. / This publication is with permission of the rights owner freely accessible due to an Alliance licence and a national licence (funded by the DFG, German Research Foundation) respectively.

Empfohlene Zitierung / Suggested Citation:

Behr, D., Bandilla, W., Kaczmirek, L., & Braun, M. (2014). Cognitive probes in web surveys: on the effect of different text box size and probing exposure on response quality. *Social Science Computer Review*, 32(4), 524-533. <https://doi.org/10.1177/0894439313485203>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Cognitive Probes in Web Surveys: On the Effect of Different Text Box Size and Probing Exposure on Response Quality

Social Science Computer Review
2014, Vol. 32(4) 524-533
© The Author(s) 2013
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0894439313485203
sscr.sagepub.com



Dorothee Behr¹, Wolfgang Bandilla¹, Lars Kaczmirek¹,
and Michael Braun¹

Abstract

In this study, we explore to what extent the visual presentation of open-ended probes, in connection with different prior probing exposure, impacts on response quality. We experiment with two text box sizes for a specific immigrant probe (*Which type of immigrants were you thinking of when you answered the question?*). On the one hand, we use a standard size equal to the other text box sizes in the survey but oversized for the specific response task. On the other hand, we use a smaller text box which fits the response task. The other probes in the survey that use the standard text box are mainly category-selection probes that ask for a reasoning for the chosen answer value. Due to randomized rotation of questions, respondents receive different numbers of category-selection probes prior to the immigrant probe, resulting in different degrees of exposure to category-selection probing prior to the immigrant probe. For the immigrant probe, we find that respondents who receive the standard text box and who have had a high exposure to category-selection probing are more likely to provide mismatching answers: The mismatch consists of not answering the specific immigrant probe but rather providing a reasoning answer as typically expected for a category-selection probe. Thus, previous experience with probing in the questionnaire can override the actual probe wording. This problem can be minimized by considering possible carryover effects of prior probes and using an appropriate survey design strategy.

Keywords

open-ended questions, probes, text box size, questionnaire design, web survey

Introduction

The visual design of web surveys can significantly impact on response quality by motivating respondents, by providing clues on what deserves particular attention, or by specifying what kind of answer

¹ GESIS—Leibniz Institute for the Social Sciences, Mannheim, Germany

Corresponding Author:

Dorothee Behr, GESIS—Leibniz Institute for the Social Sciences, P. O. Box 122155, 68072 Mannheim, Germany.
Email: dorothee.behr@gesis.org

is expected (e.g., Christian, Dillman, & Smyth, 2007; Couper, Kennedy, Conrad, & Tourangeau, 2011; Dillman, Smyth, & Christian, 2009). Visual information can include, among others, font type, position of various elements on the web page, or the size of text boxes for open-ended questions. This article focuses on the impact of text box size on responses given to online administered open-ended questions. Special emphasis will be placed on text box size in relation to previous exposure to different text box sizes.

Open-Ended Questions

Open-ended questions require respondents to provide their answers in their own words. Following a classification by Dillman, Smyth, and Christian (2009), open-ended questions can be divided into those involving (1) numerical input (date, frequency, etc.), (2) enumeration of items (names, objects, etc.), and (3) description or elaboration. These different types come with different constraints in terms of entry format or writing space. Questions asking for description or elaboration, that is, for narrative responses, come with the fewest constraints. Here, respondents are asked to write what is on their minds without format (and possibly space) restrictions. In this study, we are dealing with a special type of narrative open-ended questions, namely cognitive probes.

Cognitive Probing on the Web

Research has recently begun to incorporate cognitive probes into web surveys (e.g., Behr, Kaczmirek, Bandilla, & Braun, 2012). Cognitive probes are commonly used in face-to-face cognitive interviewing. Examples would be a comprehension probe on the meaning of a particular key term, a category-selection probe on the reasons why a certain answer category has been chosen or a specific probe addressing some more details of an item (Prüfer & Rexroth, 2005; Willis, 2005). The goal of web probing is to collect qualitative data similar to that of cognitive interviewing which allows to draw inferences on the validity of items. Web probing has the advantage that it can do without real cognitive interviewers and that it can be based on larger samples; the former allows for standardization of procedures and the latter for quantification of results. One should, however, not expect that web probing provides the same in-depth information as face-to-face cognitive interviewing since the back and forth of personal cognitive interviews is not possible on the web. When designing cognitive probes for web surveys, one should take into account the general design rules and recommendations that exist for (narrative) open-ended questions.

Visual Presentation and Web Design of Open-Ended Questions

Researchers have increasingly focused attention on the impact of visual presentation of web survey questions on response quality (see Toepoel & Couper, 2011, for a detailed listing of such research). According to Dillman et al. (2009), the four design elements words, numbers, symbols, and graphics all communicate meaning to respondents and guide them in the processing and answering of survey questions. These four elements can be modified regarding the design properties size, font, coloring, or location. In their design guidelines regarding visual presentation, Dillman et al. stress in particular the consistent presentation of similar types of information (e.g., consistency in terms of size or font styles). Respondents will then have a better chance of learning the meaning assigned to a design element or property and apply it throughout the questionnaire. This eases comprehension and helps to speed up the answering process.

The influence of visual presentation on response quality has been assessed and confirmed both for closed-ended and for open-ended questions in web surveys. For open-ended narrative questions, the size of the text box, for instance, plays an important role. Dennis, deRouvray, and Couper (2000)

find more words being written when a larger text box is used compared to a smaller one. Similarly, Smyth, Dillman, Christian, and McBride (2009) find that larger text boxes lead to an increase in response quality (measured by increased length of responses and higher number of themes), at least among late survey respondents. Respondents thus seem to infer from a large text box that a longer answer is desired (similar results for paper-and-pencil surveys can be found in Israel, 2010, for instance). It is noteworthy that a large text box can also lead to unwanted and unnecessary information when the text box is oversized and not fitting for the task. This is shown by Smyth, Dillman, and Christian (2007, cited in Dillman et al., 2009) for open-ended questions requiring an enumeration of elements. The authors compared separate small text boxes for the individual entries versus one large text box for all entries and found more unnecessary detail and elaboration in the large text box treatment than in the treatment with the individual small text boxes. Following these previous research results, we assume (1) that a text box for a narrative probe that is too large for the desired answer type leads to unnecessary and unwanted information, thus leading to reduced response quality.

In line with the call for consistent use of visual elements and properties, one should keep elements or properties fulfilling the same purpose identical. This will reduce the amount of energy that respondents need to invest to respond to a survey item, lending efficiency to the whole process of answering a survey. Approaching the issue from another perspective, one may also say that one element or property should be used for one purpose only (Dillman, Gertseva, & Mahon-Haft, 2005; Dillman et al., 2009). In our study, we are taking the example of a large text box for a narrative probe and apply it to different probe types. We assume (2) that a large text box that is consistently and regularly used throughout the survey predominantly for a certain probe type (in our study, a category-selection probe) leads respondents astray when used for a different probe (in our study, a specific probe), namely by falsely indicating to the respondents that the familiar type of answering is once more expected. Expectations due to previous exposure may override the actual probe wording in this case—the result may be a reduced overall response quality for the newly introduced probe. This assumption is linked to the idea of context effects. The innovative part of this study is that these context effects, if they occur, pertain to the perception of design rather than the substantive interpretation of an item (context effects in terms of substantive interpretation are summed up in Tourangeau, Rips, & Rasinski, 2007, pp. 197–229).

Method

The Sample

In January 2011, we conducted identical web surveys in Canada (English speaking), Denmark, Germany (eastern and western Germany as separate regions), Hungary, Spain, and the United States. The participants were drawn from the nonprobability online panels of *respondi* and international partners of the company. We targeted nationals of the respective countries aged 18–65. Quotas based on gender, age (18–30, 31–50, and 51–65), and education (lower education vs. higher education) were used to obtain a balanced, albeit not representative, sample.

The Questionnaire

The survey was designed to test the use of probing techniques in web surveys with the overall goal to assess item comparability across countries. The questionnaire covered the topics of family and politics. These two topical blocks were rotated to control for sequence effects. In total, the questionnaire contained 36 closed-ended items, with 1 item per screen. In addition, each respondent received eight open-ended cognitive probes (on a separate screen after the closed item). Soft checks were included with all items and probes, that is, after receiving a reminder in case of no response, the respondents could either answer the question or proceed without giving an answer. The closed-ended items

Standard text box, as used mainly for category-selection probes

Please explain why you selected "disagree".
 The statement was: One parent can bring up a child as well as two parents together.

Back
Continue

Condition A: Specific immigrant probe and standard text box

Which type of immigrants were you thinking of when you answered the question?
 The previous statement was: Immigrants are generally good for America's economy.

Back
Continue

Condition B: Specific immigrant probe and small text box

Which type of immigrants were you thinking of when you answered the question?
 The previous statement was: Immigrants are generally good for America's economy.

Back
Continue

Figure 1. Screenshots of a standard text box as used throughout the survey (top), a standard text box for the specific immigrant probe (Condition A), and a small text box for the specific immigrant probe (Condition B).

mainly came from the International Social Survey Programme (ISSP), among which 4 items regarding beliefs on immigrants from the ISSP 2003 questionnaire on “National Identity.” The wording was slightly adapted to make it suitable for the 1-item-per-screen design. The wording was “There are different opinions about immigrants from other countries living in [country]. By ‘immigrants’ we mean people who come to settle in [country]. How much do you agree or disagree with the following statement?” (1) *Immigrants increase crime rates.* (2) *Immigrants are generally good for [country’s] economy.* (3) *Immigrants take jobs away from people who were born in [country].* (4) *Immigrants improve society by bringing in new ideas and cultures.* A 5-point Likert-type scale from

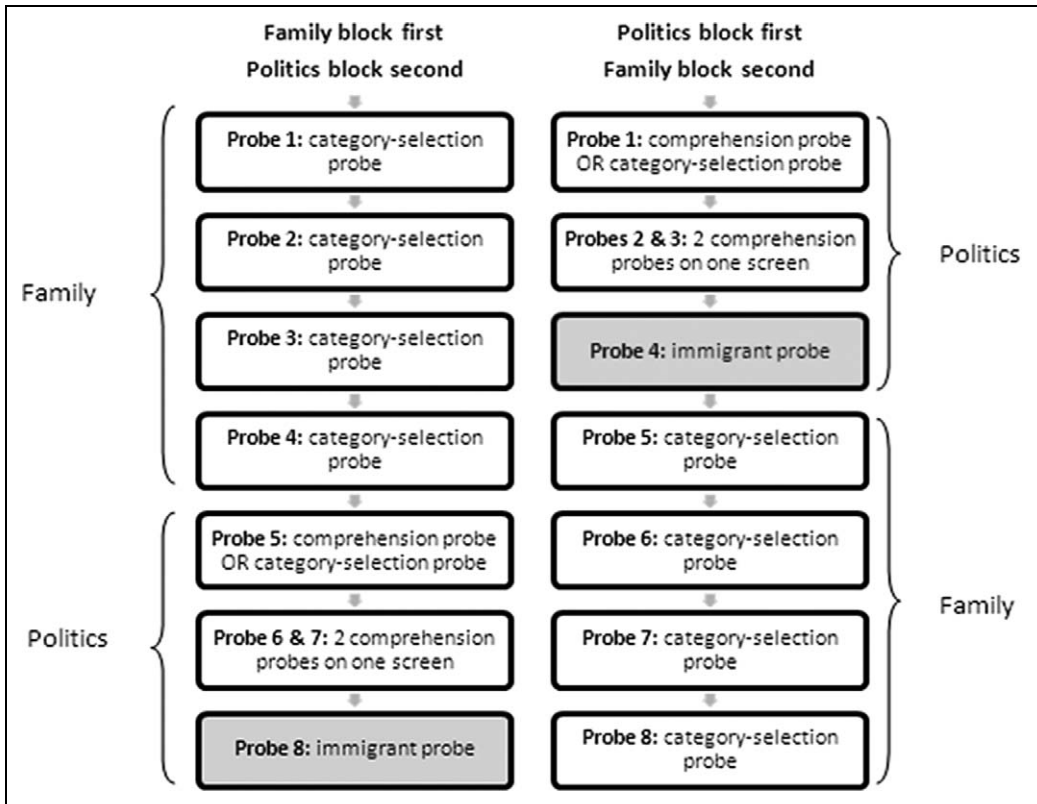


Figure 2. Questionnaire setup taking into account the random rotation of the topical blocks family and politics.

strongly agree to strongly disagree was used. Four random splits were employed in our web survey and each of the 4 items figured as the first item in one of these splits. Only the first item in each split was followed by a specific probe, asking “Which type of immigrants were you thinking of when you answered the question?” With this immigrant probe, we implemented our text box experiment. Figure 1 shows this experiment.

The top screenshot in Figure 1 shows the usual text box size, as used with the other seven probes in the survey, which were predominantly category-selection probes asking respondents for their reasons for having selected a particular scale value. The next two screenshots show what the experimental conditions regarding the immigrant probe looked like. In Condition A of the immigrant probe, the size of the text box was equal to the one used for the other probes in the survey; that is why we refer to it as *standard* text box. At the same time, this text box was oversized in view of the desired answers to the specific probe; after all, the desired naming of nationalities or ethnicities would easily fit into one line of text. In Condition B of the immigrant probe, we used a smaller text box that was more in line with the desired answers of the specific probe. This text box will subsequently be referred to as *small* text box. With both text box designs, we did not constrain writing space so that respondents were able to write more than the actual visual space indicated. Following our first assumption, the standard text box will lead to more unwanted and unnecessary information than the small text box.

Respondents received the specific immigrant probe at different times during the survey. Due to rotation of the two topical blocks family and politics, the immigrant probe was either the last of eight probes altogether or the fourth probe (see Figure 2).

Table 1. Distribution of Probe Answers for the Immigrant Probe, Separate for Each Probed Item and in Total.

Answer Code	Crime	Economy	Culture	Work	Total
Substantive answer/answers	68.8 (640)	65.8 (613)	68.3 (628)	70.7 (644)	68.4 (2,525)
Opinion statements	11.4 (106)	14.5 (135)	12.7 (117)	11.3 (103)	12.5 (461)
Nonresponse	12.6 (117)	14.5 (135)	11.3 (104)	10.1 (92)	12.1 (448)
Other answers	7.2 (67)	5.3 (48)	7.7 (71)	7.9 (72)	7.0 (258)
Total	100 (930)	100 (931)	100 (920)	100 (911)	

Note. Absolute numbers are in parentheses.

In the former position (Probe #8), respondents received four or five category-selection probes prior to the immigrant probe, which equals a high exposure to category-selection probing in connection with the standard text box. In the latter position (Probe #4), respondents received only one or no category-selection probe prior to the immigrant probe, which equals a low exposure to category-selection probing in connection with a standard text box. The variations in terms of four/five or zero/one category-selection probe/probes was due to one random probe split in our questionnaire according to which respondents received, for one particular item, either a category-selection or a comprehension probe. In focusing our descriptions on the category-selection probes, we disregard the two or three comprehension probes that *each* respondent received immediately before the immigrant probe. Since the exposure to this probe type is about equal for all respondents, we did not regard it as particularly influential in determining the results. Following the second assumption, a high prior exposure to category-selection probing affects those respondents who are in the standard text box condition of the immigrant probe. These respondents perceive the standard text box size, assign meaning to it based on previous probing experience, and eventually respond to the immigrant probe as if it were a category-selection probe.

Translation and Coding of Probe Responses

Responses to the immigrant probe were translated by professional translators from Danish, Hungarian, and Spanish into German. The German, Canadian, and American answers were not translated but coded in the original language. The translators received guidance on how to translate (and comment on) the answers to ensure translation and, subsequently, coding quality. After translation, a detailed coding scheme was set up, including, among others, codes for specific countries of origin/ethnicities (e.g., Islamic countries, Eastern Europe, or Asia) or for immigrant groups defined by a positive behavior/characteristic (*immigrants with good education*) or a negative behavior/characteristic (*immigrants who live on welfare*). The full coding scheme, intercoder reliabilities, and substantive analyses can be found in Braun, Behr, and Kaczmarek (2012). For the purposes of this article, we will present a merged coding scheme of four codes. Table 1 displays the distribution of these codes for the four probed items and in total, respectively. The merged codes are:

1. *Substantive answer/answers*: Answers referring to immigrants in general, specific nationalities/ethnicities, immigrants defined by a negative/positive behavior or characteristic, and so on.
2. *Opinion statements*: Answers stating reasons why a certain answer value has been chosen or providing a general opinion statement on immigrants rather than specifying the type of immigrants respondents had thought of. For instance, a Canadian respondent, after having answered the closed item *Immigrants improve society by bringing in new ideas and cultures*, replied to the specific probe regarding the immigrant type he or she had in mind: "Many years ago it was not such an issue and it was good to have new cultures etc, but now immigrants are trying to

change our country instead of embracing our culture and meshing both.” An American respondent, after having answered the closed item *Immigrants are generally good for America's economy*, replied to our specific probe as follows: “No, because they work often for a lot less money giving people that are born here fewer jobs.” We regarded answers of this kind as mismatching probe answers since they were not a reply to our specific probe but rather a reply to an assumed category-selection probe. As we were interested in the types of immigrants respondents thought of (particularly in terms of ethnicity), answers of the above kind, although valuable in many other instances, did not serve our measurement goal.

3. *Nonresponse*: Nonsubstantive answers, including don't knows, refusals, no text box entry, and so on.
4. *Other answers*: Other answers that could not be assigned to any of the aforementioned codes.

Data Analysis

We focused our analyses on the answer code “opinion statements.” This code was suitable to assess both whether an oversized text box led to unwanted information (first assumption) and whether an (unwanted) answer type was kept up due to expectations following previous probing exposure (second assumption). For the analyses, we pooled the four probed items (crime, economy, culture, and work) and the six countries to provide a sound quantitative basis. The binary code “opinion statements” (baseline: the remaining answers) served as the dependent variable in logistic regressions. The central independent variables in the regression models were the design variables “standard text box” (baseline: small text box) and “high exposure to category-selection probing” (variable differentiating between 4/5 and 0/1 prior category-selection probes, indicating high and low exposure, respectively; baseline: low exposure). Furthermore, “higher education” as a binary variable (baseline: lower education) was added to assess whether any cues picked up from the text box design differed by the educational level of respondents. The variables “standard text box,” “high exposure,” and “higher education” were entered first in a two- and then in three-way interaction. Moreover, a categorical variable “country” was added to control for country effects (baseline: United States).

Results

In total, 3,695 respondents across all countries completed the survey. The mean answer time across all countries was almost 17 min, and the overall break off rate was 11.4% ($n = 474$). Before presenting the logistic regression results, some descriptive results shall be presented in Table 2. The last column shows the occurrence of the code “opinion statements,” depending on previous exposure to category-selection probing and on text box size. We see that Group 1, that is, those respondents with a standard text box size for the immigrant probe and a high exposure to category-selection probing, is clearly set apart from the other groups (26% of opinion statements). We are further shedding light on this issue using a logistic regression, as shown in Table 3.

Model 1 shows a two-way interaction between standard text box design and high exposure to category-selection probes. A significant main effect is found for high exposure to category-selection probing. High exposure, therefore, contributes to an increased level of opinion writing. Furthermore, the interaction term between text box size and previous probing exposure is significant. Model 2 shows a three-way interaction between the above variables and, in addition, education. We included education since higher educated respondents might pick up visual cues differently or more quickly than lower educated respondents. We find that the educational level does not impact on opinion writing. However, through the introduction of the education variable, the main effect for high exposure disappears and the previously identified interaction effect increases. In Model 3, country

Table 2. Experimental Groups, Respective Sample Sizes, and Occurrence of Opinion Statements.

Group	Experimental Groups (Due to Rotation of Topical Blocks)	Experimental Groups (Due to Text Box Size for Immigrant Probe)	Number of Respondents	Occurrence of Opinion Statements in % (Absolute Numbers)
1	High exposure to category-selection probes (4 or 5)	Standard text box	$n = 901$	26.0 (234)
2	Low exposure to category-selection probes (0 or 1)	Small text box	$n = 916$	9.0 (82)
3	High exposure to category-selection probes (4 or 5)	Standard text box	$n = 945$	8.9 (84)
4	Low exposure to category-selection probes (0 or 1)	Small text box	$n = 933$	6.5 (61)

Note. $n_{\text{total}} = 3,695$.

Table 3. Logistic Regression Predicting the Writing of an Opinion Statement.

	Model 1	Model 2	Model 3
Standard text box (baseline: small text box)	1.378 (0.242)	1.531 (0.367)	1.561 (0.375)
High exposure to category-selection probing (baseline: low exposure)	1.413* (0.249)	1.244 (0.319)	1.257 (0.323)
Standard text box \times high exposure	2.562*** (0.573)	3.001*** (0.949)	2.917*** (0.925)
Higher education (baseline: lower education)		1.009 (0.267)	1.028 (0.273)
Standard text box \times higher education		0.792 (0.279)	0.770 (0.272)
Higher education \times high exposure		1.262 (0.446)	1.226 (0.434)
Higher education \times high exposure \times standard text box		0.760 (0.342)	0.794 (0.359)
Spain (baseline: United States)			0.579* (0.124)
Denmark			1.201 (0.222)
Hungary			1.153 (0.214)
Canada			1.064 (0.200)
Germany			0.882 (0.149)
n	3,667	3,667	3,667
Pseudo R^2	0.064	0.066	0.072

Note. Odds ratios, standard errors are in parentheses.

* $p < .05$. *** $p < .001$.

serves as a control variable. The interaction effect remains about the same, while significant country differences only become visible between Spain and the baseline United States. In a further model (not presented here), we added control variables that are potentially related to stating one's opinion about immigrants, namely, political interest and an index for xenophobia: None of the variables impacted on the findings of Model 3. Rather than pooling all country data, we also conducted separate analyses per country. This reduced the “ n ” remarkably and we could not fully replicate in these reduced regression models the findings presented above. However, the descriptive results for the occurrence of opinion statements, as presented in Table 2 (last column), in general also apply to the country level.

Discussion

We set out to explore what happens when an oversized text box is provided for an immigrant probe and when this text box size is already mentally linked in the mind of respondents to another probe type (due to a high exposure to it in the course of the survey). We found that an oversized text box itself did not necessarily result in reduced response quality, as measured by the occurrence of opinion statements. However, we obtained quite a large amount of opinion statements with the oversized standard text box when respondents had received four or five category-selection probes prior to the immigrant probe. What may have happened? A high exposure to category-selection probing means

that respondents get used to category-selection probing, that is, to giving their reasons for their chosen answer values. Furthermore, they get used to a specific design (the standard text box size) and relate this design to category-selection probing. With the standard text box occurring again for the specific immigrant probe, respondents with high previous exposure to category-selection probes may only have perceived the size of the text box and inferred from it that, once again, an explanation for their closed answer value is needed. Top-down processes in terms of expectations may have overridden the actual probe wording. The effect noticed should not be attributed to respondents reducing their efforts and trying to reach the end of the survey in the shortest time possible; after all, these respondents provide us with narrative text and this in itself requires effort and motivation. Still, respondents may feel that any lingering on a wording is not worthwhile if they can already guess—or have “learned” through prior experience—what is going on. They may thus pursue some sort of optimization of effort spending by not properly reading a given probe that seems familiar. Eye-tracking would be able to provide an answer in this respect. In addition, the prior category-selection probes in the survey followed (with only one exception) Likert-type scale items—the probed immigrant item was also a Likert-type scale item. This similarity might equally have contributed to voicing one’s opinion with the immigrant probe rather than solely naming the type of immigrant that respondents had thought of. Furthermore, it might also be possible that a certain group of respondents came to like voicing their opinion, and, if not explicitly discouraged by a small text box, they might have pursued what they like.

In sum, an oversized text box does not automatically lead to unwanted information. However, when respondents have gotten used to this “oversize” in relation to another probe type, they may carryover their experiences and expectations, which then results in unwanted information. The novelty of this finding is the context effect via expectations built up through high previous probing exposure.

Despite the clear effects of text box design and category-selection exposure on writing opinion statements, we have to issue a cautionary note: Our results are based only on one particular topic, namely immigrants. Since this topic lends itself to a hot debate, it may especially be linked to opinion writing.

In terms of future implementation of web probing, one should carefully consider text box size not just for an individual probe but in the context of all probes asked. A standard text box size for different types of probes should be avoided, especially when the different types of probes vary significantly regarding the level of elaboration that they require. Furthermore, if several different probe types are used, the sequence of probes should be critically assessed. The learning or expectation effect we witnessed may be prevented using different visual presentations for the different probe types.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by the German Research Foundation (DFG) as part of the PPSM Priority Programme on Survey Methodology (SPP 1292; project BR 908/3-1).

References

- Behr, D., Kaczmirek, L., Bandilla, W., & Braun, M. (2012). Asking probing questions in web surveys: Which factors have an impact on the quality of responses? *Social Science Computer Review*, 30, 487–498.
- Braun, M., Behr, D., & Kaczmirek, L. (2012). Assessing cross-national equivalence of measures of xenophobia: Evidence from probing in web surveys. *International Journal of Public Opinion Research*.
- Christian, L. M., Dillman, D. A., & 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.

- Couper, M. P., Kennedy, C., Conrad, F. G., & Tourangeau, R. (2011). Designing input fields for non-narrative open-ended responses in web surveys. *Journal of Official Statistics, 27*, 65–85.
- Dennis, M., deRouvray, C., & 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.
- 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.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.). New York, NY: John Wiley.
- Israel, G. D. (2010). Effects of answer space size on responses to open-ended questions in mail surveys. *Journal of Official Statistics, 26*, 271–285.
- Prüfer, P., & Rexroth, M. (2005). Kognitive Interviews [cognitive interviews]. *ZUMA How-to-Reihe 15*. Retrieved August 08, 2012, from http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/howto/How_to15PP_MR.pdf
- Smyth, J. D., Dillman, D. A., & Christian, L. M. (2007). *Improving response quality in list-style open-ended questions in web and telephone surveys*. Paper presented at the annual conference of the American Association of Public Opinion Research, Anaheim, CA.
- Smyth, J. D., Dillman, D. A., Christian, L. M., & 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.
- Toepoel, V., & Couper, M. (2011). Can verbal instructions counteract visual context effects in web surveys? *Public Opinion Quarterly, 75*, 1–18.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. New York, NY: Cambridge University Press.
- Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage.

Author Biographies

Dorothee Behr is a researcher at GESIS—Leibniz Institute for the Social Sciences, Mannheim (Germany). Her research interests include web survey design, comparability of cross-cultural questionnaires, and questionnaire translation. e-mail: dorothee.behr@gesis.org

Wolfgang Bandilla is a researcher at GESIS—Leibniz Institute for the Social Sciences, Mannheim (Germany). He has specialized in web survey methodology and mixed-mode surveys; e-mail: wolfgang.bandilla@gesis.org

Lars Kaczmirek is a researcher at GESIS—Leibniz Institute for the Social Sciences, Mannheim (Germany). His research interests include survey design, mixed mode, and online access panels; e-mail: lars.kaczmirek@gesis.org

Michael Braun is a researcher at GESIS—Leibniz Institute for the Social Sciences and adjunct professor at the University of Mannheim, Germany. His research interests include the methodology of cross-national survey research, migration, and family and gender roles; e-mail contact: michael.braun@gesis.org