Emulating interviewers in an online survey

de Leeuw, E.D.; Hox, J.J.C.M.; Scherpenzeel, A.C.

Published in:
JSM (Joint Statistical Meetings) Proceedings, Miami Beach, Florida, July 30-August 4, 2011 (Section on Survey Research Methods, ASA)

Publication date:
2011

Emulating Interviewers in an Online Survey: Experimental Manipulation of 'Do-Not-Know' over the Phone and on the Web

Edith de Leeuw[1], Joop Hox[2], Annette Scherpenzeel[3]
[1] Utrecht University, POBOX 80140, NL-3508TC Utrecht, Holland
[2] Utrecht University, POBOX 80140, NL-3508TC Utrecht, Holland
[3]Tilburg University, POBOX 90153, NL-5000LE Tilburg, Holland

Abstract
Members of a high quality, probability-based Internet panel were randomly assigned to one of two modes: (1) computer assisted telephone interview or (2) web survey. Within each mode a two-by-two experimental design was used: (1) explicit offering of do-not-know vs. no explicit do-not-know option, and (2) directly accepting a do-not-know vs. only accepting it after a friendly probe. Telephone- and web-mode differ in amount of spontaneous (without do-not-know-option) do-not-know answers and the number of prompted (with do-not-know-option) do-not-know responses. There was an interaction of mode and experimental condition; respondents offer more spontaneous do-not-know answers over the telephone. Follow-up of do-not-know answers with a probe was effective in both telephone and web. Emulating interviewers for probing in web surveys is not only possible, but also successful.

Key Words: Probing, do-not-know, item-nonresponse, uninformative response, internet-survey

1. Introduction

Despite their relative short history, web surveys have already had a profound effect on survey research (Couper and Miller, 2008). Web surveys share the advantages of paper self-administered questionnaires. For instance, Link and Mokdad (2005) found more self-reported heavy drinkers in a web survey compared to those in a telephone interview. This result remained strong and significant after adjusting for different demographic characteristics of respondents in both modes. Krauter, Presser & Tourangeau (2008) confirmed and extended these findings. Internet administration increased reporting of sensitive information amongst university alumni. Krauter et al (2008) had also access to record data and found a higher accuracy in web surveys; they report that web surveys increased both the level of reporting sensitive information and the accuracy compared to CATI with the more private self-administered telephone survey (IVR) in between. The greater openness in self-administered questionnaires, such as web surveys, is often attributed to the absence of an interviewer (de Leeuw, 1992; Tourangeau, Rips, and Rasisnki, 2000).

The absence of an interviewer also has disadvantages. Interviewers may guide a respondent through a complex questionnaire, thereby avoiding routing errors. Therefore, complex questionnaires with many routings were avoided in self-administered paper

questionnaires. Web surveys, being a form of computer-assisted questionnaires make it possible to administer very complex questionnaires in a self-administered form, thereby adding to their potential.

Another advantage of the presence of interviewers is that they may clarify questions and probe when respondents give a non-substantive answer. This is one of the main reasons why paper self-administered questionnaires produce more missing data than interviewer-administered surveys. To avoid a large number of missing data in the self-administered web mode, often providing a substantive answer is simply required (Dillman and Christian, 2005). In that case, it is impossible to continue the survey without clicking a substantive response category. This is not a very good idea as forcing respondents to answer may lead to more break-offs and guessing (Couper, 2008, p. 266). Besides, crude measures like forcing a response are not necessary, a web survey is very versatile and more sophisticated prompts may be used (see Wine, Cominole, Heuer, and Riccobono, 2006). Therefore, in this study we emulated a friendly interviewer probe and studied its effect on the amount of non-substantive (do-not-know) answers. We also investigated its effectiveness when an explicit do-not-know option was offered on the screen. Finally, the effects of our emulated interviewer probe over the Internet was compared with the effects of the same probe being delivered by a live interviewer in a telephone interview.

This enabled us to investigate the following research questions:
- Does offering an explicit do-not-know option produce more non-substantive answers?
- Do telephone interviews and web surveys differ in this respect?
-Does following up with a friendly probe help in reducing the amount of non-substantive answers?
- Do telephone interviews with a real interviewer differ from web surveys with an emulated interviewer in reducing the amount of non-substantive answers?
- Does probing negatively influence the reliability of the resulting substantive answers?

2. Method

2.1 Respondents

In this study members of the Dutch LISS panel (Longitudinal Internet Studies for the Social Sciences) were investigated. The LISS panel was established in autumn 2007 and consists of almost 8000 individuals that complete online questionnaires every month. It is based on a probability sample of households drawn from the Dutch population register by Statistics Netherlands. The sample includes people without an Internet connection. All people in the sample were approached in traditional ways (by letter, followed by telephone call and house visit) with an invitation to participate in the panel. Households that could not otherwise participate because they had no Internet access are provided with a computer and broadband Internet connection.

LISS panel members complete online questionnaires every month, for which they get an incentive of €7.50 per half hour of interview time. Each month, there can be several short or one or two longer questionnaires. It usually takes between 30 and 40 minutes to complete them all. Panel members have a personal login to the LISS panel website and can take the whole month to complete the questionnaires. They are invited by email and receive two reminders spread over the month when they do not complete the questionnaires.

We randomly assigned members of the LISS panel over mode (CATI vs CAWI): 2000 were assigned to the CATI-condition and 6134 were assigned to the CAWI (web)-condition. In the CATI-condition 1207 members of the 2000 responded (60%), and in the CAWI-condition 4003 of the 6134 responded (65%). The CAWI-fieldwork was done by the regular staff of the LISS-panel. The CATI fieldwork was done by the Dutch TNS-NIPO organization, but during the data collection a member of the LISS-research team was present as extra quality controller. For this experiment all respondents received a bonus of 6.50 euro in addition to their regular incentive payment.

2.2 Experimental Design

In spring 2009 we implemented a split ballot experiment manipulating the do-not-know-options both within a telephone interview and within a web survey. Respondents either were explicitly offered a do-not-know response category or not. When 'do-not-know' was not explicitly offered, it was still possible to express do-not-know by clicking on continue in the web survey. This to make the web-survey equivalent to the telephone survey, where it is always possible to spontaneously say 'do-not-know' to an interviewer.

All do-not-know responses were followed-up by a probe. In the telephone interview, interviewers uses a prescribed probe, saying "Thank you, I have noted down your answer. Could you perhaps state a preference for yes or no? That would be a big help to us. The question was <question was repeated>, do you lean towards yes, towards no, or do you really do-not-know?". To emulate an interviewer, the same text appeared in a pop-up box in the web survey.

Our experiment was implemented within a mode-experiment, making it possible to evaluate the influence of offering do-not-know on the number of non-informative answers for the web and compare these with the results of a telephone survey. First of all LISS-panel members were randomly assigned to a data collection mode. Within each mode respondents were again randomly assigned to the do-not-know experiment; this resulted in a two-by-two design with four experimental groups. The experimental factors were: I. data collection mode (CATI vs. web), and II. do-not-know-option (no do-not-know option vs. explicitly offering a do-not-know response category).

2.3. Questionnaire

Six questions on the use of embryo's in medical science were presented to the respondents. For example: 'Do you think that embryo's are used in scientific medical studies searching for cures to severe diseases', and 'Do you think that embryo's are used in scientific medical studies into hereditary diseases'. Response categories were 'yes' and 'no'. These questions were part of a well-tested questionnaire that was used previously to investigate the opinions of the Dutch population regarding use and acceptance of advanced medical technology. The original researchers used a self-administered questionnaire and explicitly included a do-not-know-response option. (Steegers, Dijstelbloem, and Brom, 2008).

2.4. Analyses

We calculated the number of do-not-know responses over the six questions before and after probing, thereby creating two new variables. Analyses of variance was used to investigate the effect of mode and of offering do-not-know on the number of do-not-know responses.

We also investigated whether offering do-not-know and using probes influences reliability by calculating Cronbach's reliability coefficient alpha (Cronbach, 1951) over the six questions for each experimental condition and before and after probing. Coefficient alpha gives a lower bound for the reliability (i.e., the squared correlation between observed scores and 'true' scores on a multiple item scale) and can be interpreted as the proportion 'true' score variance in the observed scores. For a thorough discussion of reliability measures in the context of survey errors, see Groves (1989, chapter 1, section 1.5). The difference in reliability between conditions was tested for significance using a meta-analytic regression approach, based on a normalizing transformation of coefficient alpha (Bonett, 2005; Hox, 2010, p. 209).

Because the response rates between the two modes differed slightly (CATI 60% and CAWI 65%) we first checked if differential nonresponse influenced the comparability of the two experimental mode conditions. For all the LISS-panel members both biographical and psychographical information is available. When we compare the respondents in the CATI-CAWI conditions, we find some small but significant differences. CATI and CAWI respondents differ in age (47.8 vs. 46.2 years, p=0.02), household size (2.9 vs. 2.8, p=0.00), (non)urbanicity (3.1 vs. 3.0, p=0.05), house-ownership (78.5% vs. 75.0%, p=0.0). In a multivariate analysis using logistic regression, only age and household size remained significant. The differences are very small, nevertheless a propensity score weight was constructed based on a logistic regression using age and household size as predictors. The mean weight is 1.00, with an standard deviation of 0.04 and a range from 0.91-1.27. All analyses were carried out on both weighted and unweighted data, since the differences are negligible, only the unweighted results are reported.

3. Results

3.1 Mode and offering an explicit do-not-know

Our first question was whether CATI and CAWI differ in the effect of an explicitly offered do-not-know-option. The answer is yes. Two-way analysis of variance showed a significant interaction effect between mode and explicitly offering a do-not-know-option ($F(1,2890)=49.03$, $p<.0001$, partial $eta^2=0.017$). When no explicit do-not-know category was offered, respondents in the telephone condition offered more spontaneous do-not-know answers before probing than respondents in the web-condition; but when an explicit do-not-know option was offered the pattern was reversed: more do-not know answers over the web.

There was also a significant main effect of offering a do-not-know option in the expected direction: more uninformative (do-not-know) answers before probing when an explicit do-not-know response category was offered ($F(1,2890)=420,14$, $p<.0001$, partial $eta^2=0.127$). There was no significant main effect of mode on number of non informative before probing ($F(1, 2890)=3.34$, $p>.068$, partial $eta^2=0.001$.

The mean number of non informative (Do-not-know) answers before probing are summarized in Table 1.

Table 1: Offering an Explicit Do-not-Know: Effect of Mode (telephone vs web) and Explicitly Offering a Do-not-know response option on number of do-know-knows.
Average number of non-informative answers over six questions *before* probing

	No DK offered	*Explicit DK offered*	*Mean Mode*
CATI	0.29	0.91	0.60
CAWI	0.05	1.32	0.69
Mean DK offered	0.17	1.12	

Mode p= .068; Dk-offered p=.000; Interaction Mode* Dk-offered p=.000; R^2= .160

3.2. Effect of Probing

3.2.1. Changes from non-informative to informative answers

Our second question was whether probing after a do-not-know response works, and if so, whether this effect differs between CATI an CAWI. In our telephone interviews each non-informative answer was followed-up with a friendly interviewer probe. In the Internet survey, we emulated the interviewer by a pop-up with the same friendly worded probe. Probing does indeed help to reduce the number of non informative answers in both CATI and CAWI. This is clearly seen in Table 2, which depicts the average number of changes from a do-not-know response to an informative answer for the experimental conditions.

Table 2: Probing and Changes from Non-informative to Informative Answer: Effect of Mode (telephone vs web) and Explicitly Offering a Do-not-know option
Average number of changes over six questions

	No DK offered	*Explicit DK offered*	*Mean Mode*
CATI	0.89	1.03	0.96
CAWI	0.71	1.41	1.06
Mean DK offered	0.17	1.12	

Mode p= .424; Dk-offered p=.001; Interaction Mode* Dk-offered p=.031<.00; R^2= .042

This reduction of non-informative answers is especially noticeable in the CAWI condition where a Do-Not-Know option is explicitly offered. This condition elicits the highest average number of DK responses (1.32, see Table1), which is considerably reduced after the probe: the average number of changes to an informative answer is 1.41 (see Table 2).

3.2.1. Number of final do-not-know-responses after probing

Probing does help to reduce the number of do-not-know answers both in the CATI and the CAWI condition. When we look into the total number of non-informative do-not-know answers after probing, we find a marginally significant main effect of mode on number of non informative after probing ($F(1, 2890)=3.77$, $p=.05$, partial $eta^2=0.001$. We also find a significant main effect of offering a do-not-know after probing ($F(1, 2890)=86.83$, $p<.0001$, partial $eta^2=0.029$), but this effect is much smaller than the effect before probing, where the partial eta^2 was 0.127. Probing does help and it helps especially when it is necessary to explicitly offer a do-not-know response option. Finally, we did find a small, but significant interaction effect between mode and explicitly

offering a do-not-know-option ($F(1, 2890)=21.42$, $p<.0001$, partial $eta^2=0.007$). The mean number of non informative (Do-not-know) answers after probing are summarized in Table 3.

Table 3: Effect of Mode (telephone vs web) and Explicitly Offering a Do-not-know response option on final number of do-know-know answers *after* Probing.
Average number of non-informative answers over six questions

	No DK offered	*Explicit DK offered*	*Mean Mode*
CATI	0.12	0.28	0.20
CAWI	0.02	0.51	0.27
Mean DK offered	0.07	0.39	

Mode p= .052; Dk-offered p=.000; Interaction Mode* Dk-offered p=.000; R^2= .046

3.4. Consequences of missing data

3.4.1. Consequences for statistical analysis

Item missing data hamper statistical analysis. A simple solution is so called listwise deletion or complete cases analysis, where all cases with one item missing are removed from the analysis. This method has some serious disadvantages. First of all, there is a very strong assumption underlying this solution, namely that the missing values are Missing Completely At Random (MCAR; Little and Rubin, 1987). Although listwise deletion is easy to implement, it can often lead to problems such as inefficient and/or biased estimates (Raesler, Rubin, and Schenker, 2008). Furthermore, in a multivariate analysis, even a small number of missings for each individual variable may add up to a substantial loss in cases and power for the multivariate analysis.

To illustrate this problem, we computed the number of cases that would be deleted using listwise deletion, when the six questions in our questionnaire would have been used in a multivariate analysis or if a composite (sum) score over the six questions would have been computed. Again we did this before (Table 4) and after (Table 5) probing.

Table 4: Consequences of listwise deletion in analysis with 6 variables.
Effect of Mode and Offering a Do-not-know response option *before* probing.
Average proportion of deleted cases

	No DK offered	*Explicit DK offered*	*Mean Mode*
CATI	0.19	0.61	0.40
CAWI	0.04	0.58	0.31
Mean DK offered	0.12	0.59	

Mode p= .000; Dk-offered p=.000; Interaction Mode* Dk-offered p=.000; R^2= .235

Even when no explicit do-not-know options is offered almost 20% of the cases would have been deleted in CATI. Of course, one could try to impute the missing values to get a complete dataset, but prevention is the better cure (De Leeuw, Hox & Huisman, 2002). As Table 5 clearly shows, a friendly probe helps reducing the number of missing values both in a telephone interview and in an internet survey, and thereby increases the number of cases available for analysis.

Table 5: Consequences of listwise deletion in analysis with 6 variables.
Effect of Mode and Offering a Do-not-know response option *after* probing.
Average proportion of cases deleted cases

	No DK offered	*Explicit DK offered*	*Mean Mode*
CATI	0.07	0.20	0.14
CAWI	0.01	0.20	0.11
Mean DK offered	0.04	0.20	

Mode p= .021; Dk-offered p=.000; Interaction Mode* Dk-offered p=.009; R^2= .054

3.4.2. Consequences for data quality

Probing does help to reduce the number of non-informative answers in telephone and web surveys. But, one may wonder if the probed substantive answers are as valid as spontaneous informative answers, and really reflect the opinions or knowledge asked for.

To investigate if probing results in more guessing and random error we calculated the reliability over the six knowledge questions before (Table 6) and after (Table 7) probing.

Table 6: Reliability (coefficient alpha) based on six questions. N=1747
Effect of Mode and Offering a Do-not-know response option *before* probing.
Cronbach's alpha and between brackets number of cases this was based on.

	No DK offered	*Explicit DK offered*	*Mean Mode*
CATI	0.447	0.558	0.482
N	(492)	(231)	(723)
CAWI	0.737	0.686	0.713
N	(549)	(475)	(1024)
Mean DK offered	0.600	0.644	)
	(1041)	(705)	

Mode p< .000; Dk-offered p<.000; Interaction Mode* Dk-offered p=.000

Before probing, we see that responses given over the telephone result in a lower reliability and more random error than responses given in a self-administered mode (web). We also see that when an explicit do-not-know option is offered, the reliability slightly increases, which may be the result of more guessing when do-not-know is not offered and the respondent is forced to choose a specific response.

However, our friendly web-probe did not reduce the reliability as Table 7 illustrates. When we compare the cells in table 6 and Table 7, we see that the values of coefficient alpha hardly change and that the same pattern occurs: higher reliability coefficients in the self-administered web mode.

Table 7: Reliability (coefficient alpha) based on six questions. N=2500
Effect of Mode and Offering a Do-not-know response option *after* probing.
Cronbach's alpha and between brackets number of cases this was based on.

	No DK offered	*Explicit DK offered*	*Mean Mode*
CATI	0.439	0.493	0.464
N	(565)	(479)	(1044)
CAWI	0.744	0.657	0.691
N	(564)	(892)	(1456)
Mean DK offered	0.591	0.600	
	(1129)	(1371	

Mode p< .00; Dk-offered p<.04 Interaction Mode* Dk-offered p=.00

4. Summary and Discussion

In a large study of the Dutch population we investigated the influence of offering a do-not-know response option in a telephone and Internet survey. We did find a significant interaction effect, when no explicit do-not-know category was offered, respondents in the telephone condition offered more non-informative (do-not-know) answers before probing than respondents in the web-condition. However, when an explicit do-not-know option was offered the pattern was reversed: more do-not know answers over the web. This confirms the hypothesis that it is easier to express a spontaneous (not-explicitly offered) do-not-know answer to a real person. This also emphasizes the importance of including all relevant response categories (e.g., Fowler and Cosenza, 2008) and pretesting these in a self-administered questionnaire (ISO, 2006).

Probing does not produce more guessing. The reliability over six questions does hardly change after probing, which indicates that probing does *not* result in more random error. We must emphasize that the probe used was a friendly probe and that respondents were *not* forced to opt for an substantive answer. In a web survey, respondents are usually confronted with an error message forcing them to give a substantive answer. In our friendly web probe we tried to emulate an interviewer and asked which answer fitted best. Among the response categories mentioned in the probe was also an explicit do-not-know. Respondents could therefore also opt for a non substantive answer after the probe and were not brutally forced to pick an arbitrary substantive answer.

We also find that both before and after probing the reliability is lower in the telephone survey than in the web survey. This is in line with earlier findings (de Leeuw, 1992, chapter 6) that self-administered questionnaires results in less random error than telephone interviews.

Our main interest was if emulating a friendly interviewer probe in a web survey would effectively reduce the number of do-not-know answers. In our telephone interviews each non-informative answer was followed-up with a friendly interviewer probe. In our Internet survey, we emulated the interviewer by a pop-up with the same friendly worded probe. Emulating a friendly interviewer probe in a web survey is indeed successful. Not only does probing help to reduce the number of non informative answers in both CATI and CAWI, it is especially effective when an explicit do-not-know-option is offered in a web survey. We therefore advice always to include a friendly probe whenever conceptual

reasons make it necessary to explicitly offer a do-not-know option in a (web) survey; for instance, in those situations when it is expected that a non-negligible number of respondents really do not have then necessary knowledge to answer.

Acknowledgements

The authors thank Stephanie Stam (CentERdata, Tilburg University) and Robbert Zandvliet (TNS-NIPO) for their valuable assistance during the data collection phase.

References

Bonett, D. G. (2002). Sample size requirements for testing and estimating coefficient alpha. *Journal of Educational and Behavioral Statistics, 27*, 335 – 340.

Couper, Mick P. (2008). *Designing Effective Web Surveys.* New York: Cambridge University Press

Couper, Mick P. and Peter V. Miller (2008). Web survey methods: Introduction. *Public Opinion Quarterly*, 17, 831-835.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.

Dillman, Don A. and Melani Christian, Leah (2005). Survey mode as a source of instability in responses across surveys. *Field Methods, 17, 1*, 30-52.

Fowler, Floyd F. and Cosenza, Carol (2008). Writing effective questions. In Edith D. de Leeuw, Joop J. Hox, and Don A. Dillman. *International Handbook of Survey Methodology.* New York: Taylor & Francis Group (Psychology press, LEA).

Groves, Robert M. (1989). *Survey errors and survey costs*. New York: Wiley.

ISO (2006) International Standard for Market, Opinion and Social Research –Vocabulary and Service Requirements. First edition 2006, Geneva: ISO.

De Leeuw, Edith D. (1992). *Data Quality in Mail, Telephone, and Face-to-Face Surveys.* Amsterdam: TT-Publikaties. Available online: www.xs4all.nl/~edithl (Retrieved September 10, 2010).

De Leeuw, Edith D, Hox, Joop, and Huisman, Mark (2005). Prevention and treatment of item nonresponse. *Journal of Official Statistics (JOS), 19*, 153-176. Also available at www.jos.nu (Retrieved September 10, 2010)

Hox, J.J. (2010). *Multilevel analysis methods*. New York: Routledge.

Krauter, Frauke, Presser, Stanley, and Tourangeau, Roger (2008). Social desirability bias in CATI, IVR, and Web Surveys. The effect of Mode and Question Sensitivity. *Public Opinion Quarterly, 2008, 72, 5*, 847-865. This issue is freely available online at http://poq.oxfordjournals.org/content/vol72/issue5/#ARTICLES (Retrieved June 21, 2010)

Little, Rod, J.A. and Rubin (Don B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Link, Michael, W., and Mokdad, Ali H (2005). Effects of survey mode on self-reports of adult alcohol consumption: A comparison of mail, web and telephone approaches. *Journal of Studies on Alcohol*, March 2005, pp239-245.

Raessler, Suzanne, Rubin, Don, and Schenker, Naftaniel (2008). Incomplete data: Diagnosis, imputation, and estimation. In Edith D. de Leeuw, Joop J. Hox, and Don A. Dillman. *International Handbook of Survey Methodology.* New York: Taylor & Francis Group (Psychology press, LEA).

Steegers, Chantal, Dijstelbloem, Huib, and Brom, Frans, W.A. (2008). *Meer dan status alleen. Burgerperspectieven op embryo-onderzoek.* [In Dutch: Assesment and points of view of citizens on technology and embryo-research]. The Hague: Rathenau-Instituut, TA rapport 0801.

Tourangeau, Roger, Rips, Lance J., & Rasisnki, Kenneth (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.