The Effects of Question Customization on the Quality of an Open-Ended Question

February 7th, 2017

Prepared by

Yukina Chen
Yukina.Chen@nebraska.gov
Data, Research and Evaluation
Nebraska Department of Education

NEBRASKA
DEPARTMENT OF EDUCATION

Table of Contents

Abstract
This study examines the effect of question wording on data quality from an open-ended question. The open-ended question used in the study is from a web-based survey – 2016 Nebraska First Year Teacher Survey. Data quality indicators including item nonresponse, response target, ineligible response, general response, and response length are examined in the study using a series of general linear regression models. It is discovered that customized question wording leads to better data quality since it produced longer responses, and respondents answering the question in customized wording produced more correctly targeted answers. Implications and directions for future research are also discussed in this paper.

Introduction
There are two common types of survey question formats: closed-ended questions and open-ended questions. Closed-ended questions refer to when respondents are provided with pre-defined response categories. Open-ended questions refer to when respondents are provided with a blank space or a box to offer their answers.

The Nebraska Department of Education (NDE) broadly asks descriptive open-ended questions from key stakeholders like teachers, students, and parents to gain valuable comments or suggestions. One of the many projects NDE conducts annually is the Nebraska First Year Teacher Survey. This web survey was originally designed to gather information from Nebraska school principals about their opinions on the effectiveness of first year teachers prepared by Nebraska post-secondary institutions. The open-ended question in this survey asks for the principals' suggestions to inform the institutions to better prepare first year teachers. However, data received from the past show that many respondents did not provide answers aligned to the question. Instead of providing comments targeting the institutions, many respondents provided comments targeting the individual teachers. Even though teachers' performance is a good proxy for the quality of training at an institution, a response directly targeting the institution is much preferred over a response aimed at specific teachers. A solution for solving this issue and improving data quality of this open-ended question was needed by NDE since the survey data is used to guide future teacher preparation action plans.

A split ballot design was implemented in this study to test the effects of question wording. The research question of this study is whether there is a difference in response quality between those who answered the open-ended question using the *customized* question wording and those who answered the open-ended question using the *general* question wording.

Background
Within open-ended question formats, there exist three types of questions: 1) descriptive open-ended questions, where respondents are asked to give detailed and in-depth information on the question, 2) numerical open-ended questions, where respondents are asked to provide a number to answer the question, and 3) list of items open-ended questions, where respondents are supposed to provide a list to the question (Dillman et al., 2014). Examples of descriptive open-ended questions are when the survey questions ask respondents about their experience shopping at a particular store, and when employee surveys ask respondents what do they like the most about their jobs. Numerical open-ended questions often ask respondents to provide a date, frequency, count, amount, or anything of a numerical value as the answer. List of items open-ended questions may have the respondents provide their class list, a list of their family members, or a list of addresses they have lived at for the past five years. The three types of open-ended question formats each have specific goals.

Open-ended question formats allow respondents to answer the question freely without limiting their responses, and they eliminate biases created by response options (Foddy, 1994; Dillman et al., 2014). This study focuses mainly on descriptive open-ended questions as they are supposed to provide the most collaborative and detailed information among all three aforementioned question types, and they are the most used open-ended question format by NDE.

Despite the advantages of open-ended questions, there are also limitations to this format. One limitation is that people often choose to skip open-ended questions. Research shows that open-ended questions generally receive higher item non-response rates in comparison to closed-ended questions (Reja at el., 2003; Frew at el., 2003; Griffith et al., 1999). One possible explanation for the higher item non-response rates is that the respondent burden is higher for open-ended questions since respondents are putting more time and effort into answering those questions. Many studies have suggested that willingness to participate in a survey is negatively associated with perceived respondent burden (Porter, 2004; Dillman et al., 2014; Peytchev, 2009).

Another limitation of open-ended questions is that respondents may provide insufficient answers to the question. For example, if a question asks, "How many days during the last week have you been home sick?" and the respondent answers "32", "32" is clearly an invalid answer, because there are only seven days in a week. With the help of technology, web surveys may embed a validation check in the question, restricting respondents to only provide a number between 0 and 7 to help reduce this type of measurement error. For descriptive open-ended questions, respondents could provide answers such as "…" or "N/A" or "I don't have a comment at this moment" that do not satisfy the needs of the question, and these comments are hard to detect until after the data have been collected.

When respondents do decide to answer, and answer in sufficient formats, they may only provide a short answer compared to the detailed and rich information that the survey researcher wants. For example, when asking employees to provide comments on what they like the most about the workplace, the survey researcher may want to get in-depth and detailed evaluations on the workplace culture, the relationship with supervisors, the relationship with co-workers, and beyond. When respondents provide short answers such as "the pay" or "the people", they do not live up to the survey researchers' initial expectations. Respondents may be motivated to provide short answers to reduce respondent burden. Dillman et al. (2014) point out that this is more likely to happen in self-administered surveys, because respondents have to now type the answers by themselves instead of have the interviewers record their answers, and respondents in self-administered surveys also do not receive motivation probes from the interviewers to provide an adequate answer.

Last, respondents may provide a detailed and sufficient answer, but the answers do not align to the survey question. For example, if a survey question asks students to provide suggestions to help create a safe campus, but the respondents produce answers related to creating more on-campus job opportunities. The suggestions provided may be detailed and valuable, but they do not answer the question directly. This measurement error happens when respondents perceive the question meaning incorrectly, it can also happen due to inappropriate question design, such as poor question wording or visual representation (Reja et al., 2003). Surveys using the web mode may be more prone to this type of measurement error because respondents tend to read the questions faster and exert less cognitive effort for web surveys (Internet Rogator, 1998).

Due to the limitations and data quality concerns of open-ended survey questions mentioned above, it is important for survey researchers to explore ways to improve data quality to open-ended questions, especially in web surveys. One of the known factors influencing survey data quality is question wording (Biemer et al., 2011; Dillman et al., 2014). Much research have been conducted to study the effects of question wording on closed-ended questions (Smith, 1987; Schuman & Presser, 1996; Borgers et al., 2004), however, little is known about the effects of question wording on open-ended questions. With both the need from NDE and the need to fill in the research gaps, this study examines the effects of question wording on data quality of an open-ended question. Specifically, this study looks at two types of question wording: version one, which is referred to as *general* question wording in this study, uses the same question stem for all teachers graduated from different institutions (e.g. "comments to inform the institution…"); version two, which is referred to as *customized* question wording, changes the question stem to include the specific name of the institution where a teacher graduated from (e.g. "comments to inform the University of Nebraska-Lincoln…").

Question wording is thought to affect the data quality in this study based on two theories: cognitive response process theory (Tourangeau, Rips, and Rasinski, 2000) and social exchange theory (Dillman et al., 2014). According to the cognitive response process theory (Tourangeau, Rips, and Rasinski, 2000), respondents go through four cognitive steps in their mind before providing a survey response. These four steps are comprehension, retrieval, judgment, and response. Question wording can affect the cognitive process from comprehension to retrieval, and thus impact the final response provided.

Comprehension happens when a respondent is trying to understand the question wording, identify question objectives, determine the type of information that will meet the questions objectives, and determine how much work it will likely take to answer the question. Data quality could suffer if the respondent perceives the question objectives incorrectly.

In the Nebraska First Year Teacher Survey, there are 28 questions prior to the open-ended question that asks the respondent to evaluate the effectiveness of the first year teacher. However, the open-ended question asks the respondent to provide suggestions to the institution. The target of interest changes in the open-ended questions from teachers to institutions. With the general question wording, "the institution" may not distinguish the change of target between the two sections, and respondents may not notice this target change on a conscious level. Thus, they may answer the question thinking the question target is still on the teacher.

Retrieval happens when the respondent keeps the question objectives in mind, and searches his/her memory for relevant information that meets the objectives of the question. The belief sampling model (Tourangeau, Rips, and Rasinski, 2000) mentions that while there is a pool of considerations stored in someone's memory, the retrieval process only retrieves a sample of items from the pool each time, depending on what information is accessible at the time of retrieval. As mentioned earlier, questions before the open-ended question pertain to teacher evaluation, thus, memories linked with the teacher are largely accessible at the time when the respondent proceeds to the next section. The customized question wording, which includes the institution name such as "the University of Nebraska – Lincoln", may stand out in the question stem more than the general question wording of "the institution" and change what is accessible to the respondent at the time. The display of a

specific institution name may help the respondent recall memories linked to the institution, rather than memories linked to the individual teacher, resulting in a difference in the retrieval process.

Judgment happens when the respondent judges the relevance of retrieved information for the question objectives and combines the retrieved information to develop a strategy for answering the survey question. Response happens when the respondent formats and edits the response to meet question objectives. Both judgment and response are largely dependent on the information retrieved from the retrieval process, thus we arrive at the same hypotheses as in the retrieval step.

Due to the differences mentioned above in the response process, the first hypothesis of the study is that the general question wording will lead to more responses targeting at individual teachers while the customized question wording will lead to more responses targeting at the institutions.

Social exchange theory (Dillman et al., 2014) believes that respondents are more likely to answer the questions and put more effort into answering survey questions if the survey researchers are able to build a trust relationship with them. One of the many ways of building the trust relationship with respondents is through personalization (Dillman et al., 2014). Implementing personalized messages has shown to have a positive effect on survey response rates (Cook, Heath and Thompson, 2000; Heerwegh and Loosveldt, 2006; Heerwegh et al., 2005). Most of the personalization found in literatures is implemented in the survey messages, and the effects of personalization are tested at the survey level. Not much is known about the effects of personalization on survey items. From the social exchange theory perspective, implementing personalization in question wording would make the respondents feel that the survey researchers are putting effort into the survey, and this impression helps with developing a trust relationship between the respondents and the survey researchers. As a result, respondents would be more likely to provide answers with good quality as an exchange of this relationship.

Therefore, it is hypothesized in this study that customized question wording will produce higher data quality. Based on the discussion earlier about the limitations of open-ended questions, item non-response rate, response length, ineligible response rate, and general response rate are used as quality indicators in this study. General response refers to when a respondent gives a comment that contains no themes, such as "this school is great" or "he is doing a good job". These comments are useful but only to a limited extent because they do not give enough information on specific themes or contents, in other words, they do not provide enough details. If the hypothesis is true that customized question wording leads to better data quality, it is expected that respondents who answered in the customized question wording condition will have higher item response rate, lower ineligible response rate, lower general response rate, and longer response length.

A confounding variable to data quality is respondent burden. Research has found that when the respondent burden is high, respondents are more likely to reduce the cognitive work, and either skip a question or provide less adequate answers (Porter, 2004; Dillman et al., 2014; Peytchev, 2009; Krosnick, 1991). A good indicator of respondent burden in this study is the number of surveys completed by each respondent. The respondent burden will be higher for someone who has to complete the same survey 10 times than for someone who only has to do it once. Therefore, this study controls for the number of surveys completed by each respondent to eliminate the effects of respondent burden on survey data quality.

Dataset

This study uses data collected from 2016 Nebraska First Year Teacher Survey. A list of first year teachers who graduated from participating Nebraska institutions was collected. The sampling frame of all principals whose school employed any of the first year teachers was then compiled. The data for both lists came from the Nebraska Student and Staff Record System (NSSRS) and the Nebraska Teacher Certification Database. If a teacher worked at multiple schools, the survey was sent to the principal of the school where the majority of the teacher's full-time equivalency (FTE) was assigned. The survey was designed using Qualtrics and distributed electronically via email. Prenotification of the survey was sent out on February 29th, 2016 to Human Resource staff, on March 2nd, 2016 to principals, and on March 14th, 2016 to institutions. The survey email invitation was also sent out on March 14th, and email reminders were sent on March 28th, April 11th, and April 20th. The survey closed on April 25th, resulting in a field time of two months. The survey was sent to everyone in the sampling frame. In total, 987 surveys were distributed and 683 were returned, resulting in a response rate of 69.20% (AAPOR RR1).

Respondents were asked to first rate the extent to which the first year teacher was effectively prepared for their school assignment on various indicators. One example of the indicators is "Standard 3.2 - The teacher creates environments that encourage positive social interaction, active engagement in learning, and self-motivation." These indicators were based on the degree to which the teacher meets the expectations: Consistent, Frequent, Occasional, or Rare. Respondents were then asked if they considered the teacher effectively prepared for continuing employment in their districts and to provide comments "which can inform the institution's continuing improvement efforts toward preparing classroom-ready teachers."

Methods

Research Design

A split ballot design was used to test the difference in data quality between those who were assigned to the customized question wording and those who were assigned to the general question wording for the open-ended question in the Nebraska First Year Teacher Survey. Specifically, the general question wording was "Comments to inform the institution that prepared this teacher with its continuing improvement efforts toward preparing classroom-ready teachers" while the customized question wording changed "the institution that prepared this teacher" to the specific university or college name. For example, if a teacher graduated from the University of Nebraska – Lincoln, the principal who received the customized question wording version would see the question as "Comments to inform the University of Nebraska – Lincoln with its continuing improvement efforts toward preparing classroom-ready teachers".

Respondents were assigned to one of the two question wording conditions randomly. This assignment was performed in Excel by randomly assigning respondents with the number zero or one; those who were assigned with zero received the general question wording, and those who were assigned with the number one received the customized question wording.

Response Coding

Survey comments were coded at an individual level. One graduate assistant was assigned to conduct the coding, and each individual response was coded on: whether the respondent answered the open-ended question, whether the response provided was eligible, whether the response given was too

general, whether the response aimed at the individual teacher or the institution, the length of the response, and the theme(s) included in the response.

Those who answered the open-ended question received a 1, and those who did not answer received a 0 on the variable. Comments like "None", "N/A", and "No comments" counted as ineligible responses, and received a 1 (compared to 0) on the variable. General comments like "she is a great teacher" and "the institution prepares excellent teachers" did not receive any theme coding, but was coded 1 (compared to 0) on the general response variable. Whether the response aimed at the individual teacher was coded as 1 indicating it was targeting at an individual teacher and 0 indicating it was targeting at the institution. For example, "Andrew does a good job in the classroom. I am pleased with his ability" was coded 1 for this variable, while "Continue to group in classroom management and diversity training" was coded 0 for this variable. The length of the response was coded based on the character count of the comments; for example, the comment "Behavior management needs to be improved" was coded 40 for length. The theme(s) of the response was developed based on the meaning of the response, and themes were extracted from comments, for example, "Teachers need more familiarity with student data and the processes available to use/implement this information in the classroom. Most districts are using NWEA MAPS testing. At a minimum, give some exposure to this testing system and how to interpret and use the results" was coded as "data using". Note that one comment may contain multiple themes.

Data Analysis

The independent variable in this study is question wording, which was coded at two levels: 0 = general question wording and 1 = customized question wording. The dependent variables in this study are the data quality indicators: whether the respondent answered the open-ended question, whether the response provided was eligible, whether the response given was too general, whether the response aimed at the individual teacher, and the length of the response.

The control variable in this study is the number of surveys completed by each respondent. This was selected as the control variable due to a larger respondent burden for those who completed more than one survey, and some of the variance in data quality can be eliminated by controlling for the respondent burden caused by having to fill out more than one survey.

For the analysis, this study used logistic regression models to study the effects of question wording on item nonresponse, ineligible response, general response and response targeting, because all these outcome variables are binary. A linear regression model was used to study the effect of question wording on response length due to the outcome variable being continuous. The control variable was then added to each model to further examine the effect of question wording in each model.

Results

As shown in Table 1, from the returned surveys, there were 340 respondents who received the general question wording and 343 respondents who received the customized question wording. Within those 683 respondents, 323 (47%) provided an answer to the open-ended question. Of these 323 answers, 20 (6%) were ineligible responses, 59 (18%) were general responses, 167 (52%) responses targeted at individual teachers, and 131 (41%) responses targeted at institutions. The number of surveys completed per respondent ranged from 1 to 9, the average number of surveys completed per respondent was 2.37, and the average response length was 154.17 characters.

Table 1. Unweighted descriptive statistics of measures

Variables	N	Percentages (%)
Question Wording		
General	340	49.78
Customized	343	50.22
Responded	323	47.29
Ineligible Response	20	6.19
General Response	59	19.41
Response Target		
Individual	167	54.93
Institution	131	45.07
	Mean	Standard Deviation
Number of surveys completed	2.37	1.69
Response Length	154.17	123.51

The simple logistic regression model examining the difference in response target indicated that there was a difference in response target between those who answered in the general question wording and those who answered in the customized question wording. As predicted, those who answered in the customized question wording were less likely to provide responses targeting at individual teachers (OR=0.53). When holding the number of surveys completed per respondent constant, those who answered in the customized question wording were still less likely to provide responses targeting at individual teachers (OR=0.55). Table 2 presents the coefficients and odds ratios for the two models.

Table 2. Regression analyses on response target

	Model 1			Model 2		
	Coef.	SE	Odds Ratio	Coef.	SE	Odds Ratio
Intercept	-0.20	0.12		-0.49*	0.22	
Condition (0=general question wording)	-0.32**	0.12	0.53	-0.30*	0.12	0.55
Number of Surveys Completed				0.14	0.09	1.15
Model Fit Statistics						
n	304			304		
df	1			2		
Likelihood Ratio x^2	7.50			9.94		
p	0.006			0.007		

**p<0.05, **p<0.01*

According to the simple logistic regression model using question wording as a predictor to item response, no difference was found in item response between those who answered in the general question wording and those who answered in the customized question wording,. When holding the question wording constant, those who completed more surveys showed a higher item response pattern (OR=1.18), meaning those who answered more surveys are more likely to answer the open-ended question. Table 3 presents the coefficients and odds ratios for the two models.

Table 3. Regression analyses on item response

	Model 1			Model 2		
	Coef.	SE	Odds Ratio	Coef.	SE	Odds Ratio
Intercept	0.11	0.08		-0.28*	0.14	
	Model 1			Model 2		
Condition (0=general question wording)	-0.05	0.08	0.91	-0.03	0.08	0.94
Number of Surveys Completed				0.17**	0.05	1.18
Model Fit Statistics						
n	683			683		
df	1			2		
Likelihood Ratio x^2	0.42			12.92		
p	0.52			0.001		
*$p<0.05$, **$p<0.01$*						

The simple logistic regression model examining the difference in ineligible response using question wording as a predictor indicated that there was no difference in ineligible response between those who answered in the general question wording and those who answered in the customized question wording. The multiple logistic regression model including the number of surveys completed as a control variable showed the same result. Table 4 presents the coefficients and odds ratios for the two models.

Table 4. Regression analyses on ineligible response

	Model 1			Model 2		
	Coef.	SE	Odds Ratio	Coef.	SE	Odds Ratio
Intercept	2.72**	0.23		3.00**	0.43	
Condition (0=general question wording)	-0.08	0.23	0.85	-0.10	0.23	0.81
Number of Surveys Completed				-0.12	0.16	0.88
Model Fit Statistics						

	Model 1	Model 2
n	323	323
df	1	2
Likelihood Ratio x^2	0.13	0.72
p	0.72	0.70
**p<0.05, **p<0.01*		

No difference was found in general response between those who answered in the general question wording and those who answered in the customized question wording according to the simple logistic regression model. When holding the number of surveys completed per person constant, there was still no effect of question wording on general response. Table 5 presents the coefficients and odds ratios for the two models.

Table 5. Regression analyses on general response

	Model 1			Model 2		
	Coef.	SE	Odds Ratio	Coef.	SE	Odds Ratio
Intercept	1.43**	0.15		1.51**	0.27	
Condition (0=general question wording)	0.15	0.15	1.30	0.12	0.15	1.28
Number of Surveys Completed				-0.04	0.11	0.96
Model Fit Statistics						
n	304			304		
df	1			2		
Likelihood Ratio x^2	0.80			0.93		
p	0.37			0.63		
**p<0.05, **p<0.01*						

For those who answered in the general question wording, the average response length was 113.53 characters. There was no difference in response length between those who answered in the general question wording and those who answered in the customized question wording using only question wording as a predictor. When holding the number of surveys completed per person constant, the response length for those who answered in the customized question wording was longer than those who answered in the general question wording, as predicted. The difference was about 32 characters. When holding the question wording constant, those who answered more surveys had shorter response length; with one more survey completed per person, the average response length decreased by about 16 characters. Table 6 presents the coefficients for the two models.

Table 6. Regression analyses on response length

	Model 1		Model 2	
	Coef.	SE	Coef.	SE
Intercept	113.53**	22.17	140.41**	23.44
Condition (0=general question wording)	27.27	14.11	32.54*	14.00
Number of Surveys Completed			-16.45**	5.20
Model Fit Statistics				
n	304		304	
df	(1, 302)		(2, 301)	
F-test	3.74		6.92	
p	0.054		0.001	
Adjusted R square	0.0122		0.038	
**p<0.05, **p<0.01*				

Discussion and Conclusion

Based on the results above, it is concluded that the customized question wording produced higher data quality than the general question wording. Even though no difference was found in item response, ineligible response, and general response, it was found that the customized question wording was associated with longer response length and correctly targeted responses. Incorrectly targeted responses was the biggest concern of the open-ended question data quality in the Nebraska First Year Teacher Survey. Switching the general question wording to the customized question wording reduced the odds of producing incorrectly targeted responses by almost 50%. Therefore, it is suggested that the customized question wording be used in all future iterations of the Nebraska First Year Teacher Survey.

Interestingly, it was found that, when holding the question wording constant, those who completed more surveys showed a higher item response pattern. The reason why this regression coefficient was significant may be that there were more people who completed more than one survey in the dataset, and the patterns of answering the survey within one person (i.e. respond to the open-ended question) triggered the significant effect. In other words, those who answered the survey more than once and provided answers to the open-ended question led to the significant coefficient, their contribution biased the coefficient in a positive direction. The effect of respondent burden was also discovered from the results, specifically, the more surveys one needed to complete, the shorter the response length is. This agrees with the hypothesis that respondent burden needs to be considered when examining factors that impact data quality.

There are also some limitations to the study. The Nebraska First Year Teacher Survey data were collected online; therefore, this experiment needs to be replicated in other data collection modes such as interviewer-administered mode or paper-and-pencil mode to test if the same effects hold. Another limitation of the study is that the settings in this experiment were made very specific to the

Nebraska First Year Teacher Survey, thus calling for the need to duplicate this experiment in other survey settings to see if the same effects hold.

In conclusion, this study found that question customization could impact open-ended question data quality. Thus, it is recommended that the question wording be changed from general wording to customized wording in the Nebraska First Year Teacher Survey. Future research still needs to be conducted to study other factors that can influence the data quality of open-ended questions to fill in the current research gaps.

References

Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A., & Sudman, S. (Eds.). (2011). *Measurement errors in surveys* (Vol. 173). John Wiley & Sons.

Borgers, N., Sikkel, D., & Hox, J. (2004). Response effects in surveys on children and adolescents: The effect of number of response options, negative wording, and neutral mid-point. *Quality and Quantity*, *38*(1), 17-33.

Cook, C., Heath, F., & Thompson, R. L. (2000). A meta-analysis of response rates in web-or internet-based surveys. *Educational and psychological measurement*, *60*(6), 821-836。

DeMaio, T. J. (1984). Social Desirability and Survey. *Surveying subjective phenomena*, *2*, 257.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: the tailored design method*. John Wiley & Sons.

Frew, E. J., Whynes, D. K., & Wolstenholme, J. L. (2003). Eliciting willingness to pay: comparing closed-ended with open-ended and payment scale formats. *Medical Decision Making*, *23*(2), 150-159.

Foddy, W. (1994). *Constructing questions for interviews and questionnaires: Theory and practice in social research*. Cambridge university press.

Griffith, L. E., Cook, D. J., Guyatt, G. H., & Charles, C. A. (1999). Comparison of open and closed questionnaire formats in obtaining demographic information from Canadian general internists. *Journal of clinical epidemiology*, *52*(10), 997-1005.

Heerwegh, D., & Loosveldt, G. (2006). An experimental study on the effects of personalization, survey length statements, progress indicators, and survey sponsor logos in Web Surveys. *Journal of Official Statistics*, *22*(2), 191.

Heerwegh, D., Vanhove, T., Matthijs, K., & Loosveldt, G. (2005). The effect of personalization on response rates and data quality in web surveys. *International Journal of Social Research Methodology*, *8*(2), 85-99.

Holland, J. L., & Christian, L. M. (2009). The influence of topic interest and interactive probing on responses to open-ended questions in web surveys. *Social Science Computer Review*, *27*(2), 196-212.

Internet Rogator (1998): Internet Rogator Help with Surveying. Retrieved from http://www.internet-rogator.com/htm/help.htm

Peytchev, A. (2009). Survey breakoff. *Public Opinion Quarterly*, *73*(1), 74-97.

Porter, S. R. (2004). Raising response rates: what works?. *New directions for institutional research*, *2004*(121), 5-21.

Sharp, L. M., & Frankel, J. (1983). Respondent burden: A test of some common assumptions. *Public Opinion Quarterly*, *47*(1), 36-53.

Smith, T. W. (1987). That which we call welfare by any other name would smell sweeter an analysis of the impact of question wording on response patterns. *Public Opinion Quarterly*, *51*(1), 75-83.

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*, 029.

Schuman, H., & Presser, S. (1996). *Questions and answers in attitude surveys: Experiments on question form, wording, and context*. Sage.

Reja, U., Manfreda, K. L., Hlebec, V., & Vehovar, V. (2003). Open-ended vs. close-ended questions in web questionnaires. *Developments in applied statistics*, *19*, 159-177.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press.