Journal of Official Statistics, Vol. 26, No. 2, 2010, pp. 271–285

Effects of Answer Space Size on Responses to Open-ended Questions in Mail Surveys

Glenn D. Israel[1]

Open-ended questions can provide valuable information to help researchers understand a respondent's thinking but using such questions has proven problematic in mail surveys. Experimental data are used to assess the impact of answer space size on the length and content of responses to two open-ended questions that are part of Florida Cooperative Extension's customer satisfaction survey. Based on 2,200 responses collected from 2003 to 2006, the size of the answer space showed no effect on the propensity to answer, but did affect the length of the response, with larger spaces eliciting more words and themes. In addition, larger spaces were more likely to have details elaborated. These results suggest that it is better to design larger answer spaces to more fully capture high-quality responses.

Key words: Visual design; question structure; response quality.

1. Introduction

Policy makers, organizational leaders, and scientists depend on mail surveys to obtain information on many topics. The usefulness of surveys depends upon minimizing error and obtaining "optimal" responses. Optimal responses are more likely to occur when the researcher designs an instrument to communicate expectations for responding, as well as when the respondent has been diligent in retrieving and reporting relevant information (de Leeuw 2008). One dilemma facing survey designers concerns the use of open-ended questions. Open-ended questions have a reputation for poor-quality responses, difficulty in coding, and costly administration (Dillman 2007; Smyth, Dillman, Christian, and McBride 2009). In addition, survey designers need to fit items within the constraint of a page-limit so open-ended questions are often either reduced in allocated space or eliminated outright. On the other hand, open-ended questions have the potential to generate rich, detailed answers that can help researchers understand respondents' thinking and behavior. Thus, there is a need to examine how and when open-ended questions can be used in mail surveys.

Recent studies provide considerable evidence that survey results are influenced by visual design components (numbers, symbols, and graphics) as well as the words in

[1] University of Florida, P.O. Box 110540, Gainesville, FL 32611-0540, U.S.A. Email: gdisrael@ufl.edu
Acknowledgments: Revision of a paper presented at the annual Joint Statistical Meeting of the American Statistical Association on July 29, 2007, in Salt Lake City, UT. This research is part of FLA-AEC-04193, Reduction of Error in Rural and Agricultural Surveys, of the Florida Agricultural Experiment Station, University of Florida, in Gainesville, FL. The author thanks Dilcia Toro and Kat Israel for assistance in coding the content of the open-ended questions and Don Dillman and anonymous reviewers for constructive comments on an earlier draft of the manuscript.

question stems and response options. Visual factors include placement and size of arrows for navigating among questions (Redline, Dillman, Dajani, and Scaggs 2003), the location of symbols for answering questions about dates on which an event occurred (Smyth, Dillman, Christian, and Stern 2004; Christian, Dillman, and Smyth 2007), and the use of lines between and spacing of response options (Tourangeau, Couper, and Conrad 2004). Despite the volume of research on questionnaire design (see reviews by Schaeffer and Presser 2003; Tourangeau 2004; as well as Dillman 2007; Schuman and Presser 1981), information on the effect of visual design on answers to open-ended questions is limited. Using a paper questionnaire with white answer spaces on a blue background, Christian and Dillman (2004) found that the larger of the two answer spaces resulted in longer answers (measured by words) and more themes in the answers provided by college students. Similarly, Stern, Dillman, and Smyth (2007) found that respondents wrote more words in the larger space ($p = .069$) for a survey on white paper with the answer space outlined by a black line. More recently, Smyth et al. (2009) found that answer space size affected responses to a web survey but it was limited to late responders in this mode.

Though these studies provide evidence that visual design affects how people respond to open-ended questions, the studies have several limitations. First, Christian and Dillman's (2004) and Smyth et al.'s (2009) findings are based on samples of college students, so the evidence about their application to the general public or other populations is limited. This article uses a sample comprising adults who are users of a nonformal educational program available to the general public to address the first issue. The sample includes adults ranging in age from 18 to 98 (mean = 56.1 years) and education from less than high school (3.1%) to completing a graduate or professional degree (16.7%). Second, all of the previous experiments were based on comparing a large box with a small one and this test does not address whether there is an optimal size for the answer space. I test an array of box sizes in two experiments with four replications each to examine how the size of the answer space affects the length and content of answers provided by respondents to a self-administered mail survey.

2. Theoretical Background

The influence of visual and verbal languages used in self-administered surveys is intimately interwoven with the cognitive processing employed by respondents in answering questions. There is a general consensus on the cognitive process employed by respondents (Dillman 2007; Krosnick, Narayan, and Smith 1996; Krosnick 1999; Schaeffer and Presser 2003). This process is based on four steps: interpreting the question, retrieving information from memory, integrating the information into an answer, and creating the response on the survey form (Tourangeau and Rasinski 1988). In the case of open-ended questions, the size of the answer space conveys the researcher's expectation about the length of the desired answer to the respondent, which then becomes part of the calculus in interpreting the question and formulating a response. Given that respondents follow culturally prescribed rules for processing information (Tourangeau et al. 2004), an answer box that indicates space for writing four lines of text should tell a respondent to give a longer answer than one that accommodates 1, 2, or 3 lines, respectively. Following Tourangeau et al.'s (2004) framework, a white answer space on a shaded background also

contains the cultural message to write in the white area only. The figure-ground contrast also employs the "power of constraint" to encourage people to write their answers in the prescribed space (Dillman, Gertseva, and Mahon-Haft 2005). Similar graphical formulations, such as black lines on white paper to outline the answer space, should convey a similar message, while other formulations, such as a series of horizontal lines, might suggest a different expectation for responding (i.e., provide a list of items instead of description with complete sentences), and these might affect response quality in unintended ways.

Some researchers suggest also that answers to survey questions are often based on the most accessible information rather than all of the information available in memory (Schwarz, Bless, Wänke, and Winkielman 2003). This contributes, in part, to respondents' providing "satisfactory" answers instead of "optimal" ones for many questions included in surveys (Krosnick et al. 1996; Krosnick 1999). Consequently, there might be ceiling effects, where a threshold is reached and larger answer boxes do not elicit longer answers, because respondents cannot or will not report more information. By extension, people might also be discouraged from responding to an open-ended question when the size of the space indicates the need to write a lengthy, detailed answer.

It is also likely that respondent motivation, information accessibility, and processing fluency affect the quality of responses to open-ended questions (Schwarz et al. 2003). More motivated individuals, for example, might think longer and more deeply about their response than less motivated respondents and this could minimize the effect of answer space size on open-ended questions (Bishop, Hippler, Schwarz, and Strack 1988). Persons with more formal educational and those under 60 years of age also may have greater processing fluency for responding to open-ended questions, which can be viewed as being both physically (because a person writes the response) and cognitively demanding relative to close-ended questions (Stern et al. 2007). There is some evidence which suggests visual design can mediate motivation and processing fluency for some respondents, with less educated persons and men being most affected (Stern et al. 2007).

In sum, the literature gives reason to believe that the design of larger answer spaces for open-ended questions in mail surveys is likely to increase the length and improve the quality of responses, at least up to a point. These expectations are tested below.

3. Procedures

The experiments were included in an annual mail survey of adults who were participants in nonformal educational programs conducted by the Florida Cooperative Extension Service (FCES) from 2003 to 2006. FCES provides an array of educational programs, including homeowner landscape maintenance, parenting and financial management, and agricultural production classes. Although Florida's adult population comprises the potential audience for FCES programs, a comparison of client characteristics with American Communities Survey data shows participation is higher among older persons (31.7% of clients and 21.7% of residents, respectively, are 65 or older), they have more educational attainment (40.1% and 25.2%, respectively, have completed a bachelor's degree), and are more likely to be female (54.0% and 51.5%, respectively).

The self-administered survey is sent to a sample of clients who are selected from the population that have attended a workshop or seminar, called the Extension office, or visited the office in order to solicit feedback about their experiences. An equal-sized stratified random sampling design is used for selecting individuals from lists obtained from the county Extension office. With 67 counties, 13 or 14 counties participate each year so that a county contributes data every five years. The sample for each county also is stratified by type of participation (planned program versus office visit/telephone call). The intended sample size per county was 45 in 2003, 90 in 2004, 110 in 2005, and 100 in 2006 (some smaller counties collected fewer names during the recording period each year and some counties were exempted from the survey in 2004 due to hurricanes).

The survey was implemented using a sequence of contacts – preletter, survey and cover letter, reminder postcard, and second survey and cover letter to nonrespondents. Each client was randomly assigned to one of the four treatment groups and received the corresponding instrument, form A, B, C, or D, respectively. A total of 2,257 useable surveys were pooled for the analysis (275 for 2003, 442 for 2004, 747 for 2005, and 793 for 2006). The response rate (RR1, American Association for Public Opinion Research 2004) was 58.2% in 2003, 51.0% in 2004, 63.6% in 2005, and 63.3% in 2006. Despite interruptions in sequencing the mailing because of four hurricanes in 2004, response rates in affected areas were not reduced. The lowest response rate (33.5%) was in Miami-Dade County and this appeared due to proportionately fewer Hispanics returning the survey, while the remaining areas averaged 56.2%.

The two-page instrument included items on the client's experience in obtaining information, using and sharing the information, and overall rating of the service provided. A set of demographic items also was included for assessing differences based on respondent characteristics.

The experiments involved graphical manipulation of the text boxes used for the two open-ended questions in the survey. The first item, "Please explain why it did or didn't solve the problem or answer your question" was a follow-up to questions on whether the respondent had an opportunity to use information from Extension and whether the information had solved the problem or answered the question. A few individuals ($n = 27$) did not check an answer to the screening question and answered the open-ended question. Others ($n = 69$) ignored the skip instructions in the screening question and provided a response to the open-ended question. Additional analysis (data not shown) indicated that including both groups did not affect the substantive results, so all of the respondents who wrote an answer to the open-ended question were included in the analysis. The second item applied to all respondents and asked "What could we do to improve our service to you and others in the county?" Each of these items included an answer space designated by a white box on a light gray background (Figure 1).

Given that Christian and Dillman (2004) found that a larger answer space led respondents to write more words and include more themes than did a smaller answer space for three questions on a mail survey of college students, I examined an array of the answer box heights. I hypothesized that the box height suggests the number of line of text expected, as well as the volume of words for an answer. To test this idea, two heights (.68 and 1.12 inches) were used during the 2003 survey to replicate the Christian and Dillman study and four (.28, .56, .84, and 1.12 inches) for 2004 to examine the linear effect

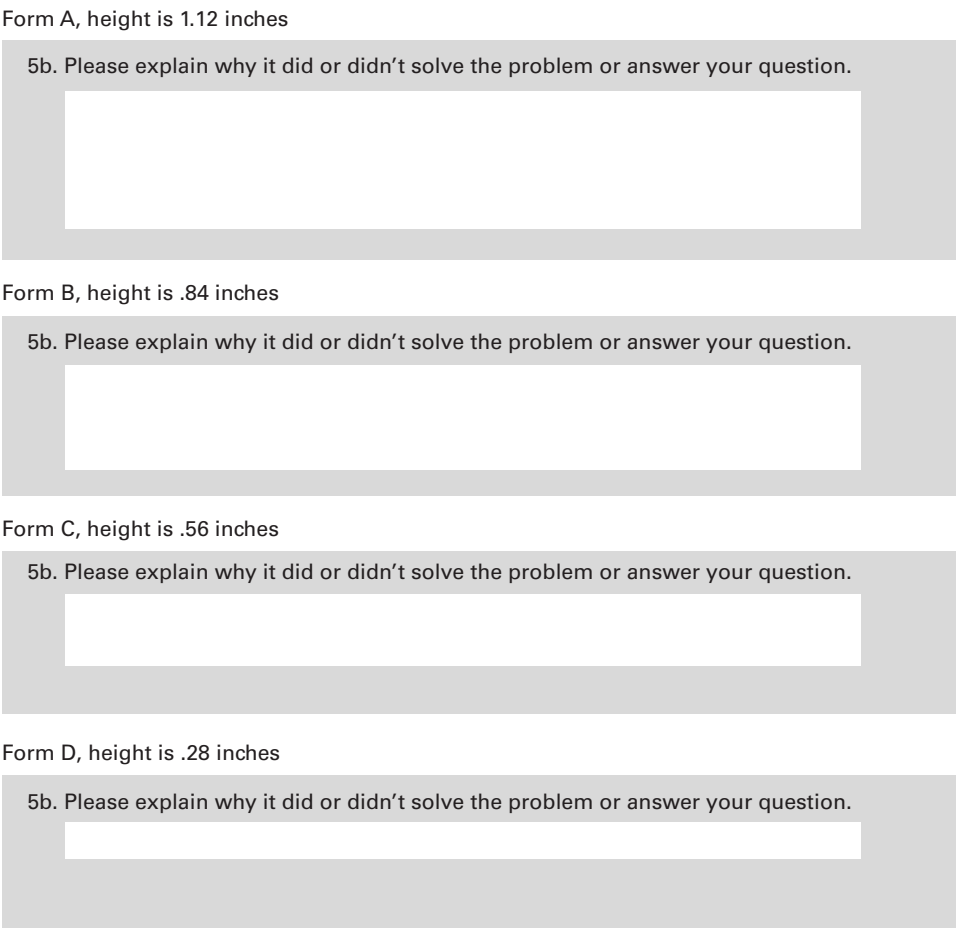

Fig. 1. Question format used to test hypothesis about graphical manipulations involving answer space, 2004 version

of increased size. Note that .28, .56, .84, and 1.12 inches provided room to comfortably write 1, 2, 3, and 4 lines of text, respectively. For 2005 and 2006, box heights were increased to test for a ceiling effect (.28, .84, 1.40, and 1.68 inches for the first question and .28, .56, 1.12, and 1.40 for the second question). Table 1 shows the combination of treatments and unit responses by year. I also hypothesized that the smallest answer space will prove inadequate for more respondents and result in a larger percentage of respondents who override the cultural norm and write outside of the designated space.

The responses to the experimental treatments were measured in two ways: the physical characteristics of the responses and the content of the response. The former includes the propensity to respond (given that an answer is not necessary for all respondents, I do not use the term "item nonresponse"), the propensity to write outside the designated answer space, and the length of the response (as measured by the number of words, lines of text, and sentences). The content was coded to count the number of themes, identify the primary content category, and assess whether additional details were elaborated. Following Smyth et al. (2009), a theme was defined as "a concept or subject that answered the

Table 1. Number of unit respondents by experimental treatment and year

	Item 1 Box height in inches						
	.28	.56	.68	.84	1.12	1.40	1.68
2003			131		144		
2004	111	112		106	113		
2005	183			180		195	189
2006	199			198		202	194
	Item 2 Box Height in Inches						
	.28	.56	.68	.84	1.12	1.40	1.68
2003		144		131			
2004	113	106		112	111		
2005	183	195			189	180	
2006	199	202			194	198	

question and was independent of all other concepts within the response." The coding was done using the open-coding method (Strauss and Corbin 1990) to develop the coding rules using three independent coders. Once the coding rules were established, two independent coders coded each answer. Where differences in coding occurred, these were discussed until a consensus was reached (in a few instances, a third coder participated in building the consensus).

The content category for the item "Please explain why it did or didn't solve the problem or answer your question" used the following codes: Positive = Extension Service was able to provide a solution to a problem or an answer to a question for a client; If the information was used or applied; Negative = Extension Service was unable to provide a solution to a problem or answer to a question; Don't know = Client doesn't know if Extension Service was able to provide him/her with a solution to a problem or answer to a question; and Other = Client expresses different ideas or points of view other than when answering whether Extension Service was able to provide him/her with a solution to a problem or answer to a question. The content category for the question "What could we do to improve our service to you and others in the county?" was coded similarly: Positive = Extension Service can be improved; Negative = Extension Service can not be improved; Don't know = Client does not know if Extension Service can be improved; and Other = Client expresses different ideas or points of view other than when answering whether Extension Service can be improved.

Finally, the presence of elaboration was coded. Following Smyth et al.'s (2009) coding of their elaboration variable, the "details" variable was coded as 1 when there were added phrases (on the situation, problem, action taken, or results for the first item and improvements for the second question) and it was coded as 0 otherwise.

The analysis of the number of words used generalized poisson regression (GPR) to estimate the effect of increasing answer space size. GPR is appropriate for count data where the dependent variable is over-dispersed (Gardner, Mulvey, and Shaw 1995; Long 1997). This method incorporates a dispersion parameter (a) to allow for correct inferences (Gardner, Mulvey, and Shaw 1995; Wang and Famoye 1997). Thus for GPR, the dependent variable, Y_i (i.e., number of words), has an expected value of u_i and follows a poisson distribution. The logit link function is used to linearize the expected value of y_i, such that $\ln(u_i) = \sum_j \beta_j x_{ij}$, where β is a vector of j regression parameters and x_i is a $j - 1$

vector of explanatory variables, including answer space size and respondent attributes. The mean and variance of Y_i are: $E(Y_i|x_i) = u_i$ and $V(Y_i|x_i) = u_i(1 + au_i)^2$, respectively.

For response probability, writing outside the answer space, and whether additional details were provided, Chi-square tests and binary logistic regression were used. Because the number of themes ranged from 1 to 4 for both of the open-ended items, Chi-square tests and ordinal logistic regression were used to estimate the effect of the answer space size. Model fit for poisson and logistic regressions was based on model Chi-square and Deviance statistics (see Cameron and Windmeijer 1996; Mittlböck and Schemper 1996, 1999; Mittlböck and Waldhör 2000).

4. Results

The data for the two experiments involving the height of the answer space showed strong effects on the extent of respondents' answers but not on the likelihood of their answering the question. The box height had no effect on the propensity to provide an answer, as shown by the *Response probability* in the top panel of Tables 2 and 3. There is no consistent trend across the size categories for either question. The *p*-values for the Chi-square test were not significant for any of the replications (except for the 2003 version of the improvement question in Table 3, which showed an increase in the response probability for the larger answer box). The pooled data also showed no significant difference for the two items.

Respondents also showed a strong tendency to write outside the answer space when presented with a shorter box height, as shown in the second panel in Tables 2 and 3. The effect is concentrated in the shortest box, at .28 inches, and suggests that this space was inadequate. The *p*-values for the Chi-square test were significant for every replication except the initial experiment in 2003 for the question about how to improve service. It is also possible that the amount of shaded space between the answer box and the next question might have encouraged some respondents to write outside the designated answer area (see Figure 1, Form D).

There was a strong association between the answer space height and the length of the response. The mean number of words panels in Tables 2 and 3 show that the number of words increased with answer box height. The initial comparisons for 2003, involving more limited comparison regarding the size of the answer space, were all in the expected direction but did not achieve statistical significance for one item. The subsequent replications in 2004, 2005, and 2006 showed highly significant effects on this physical aspect. The pooled data for mean words in Table 2 show a monotonic increase with a change in the answer box height. For the pooled data in Table 3, there is an increase in mean words with each change in height through 1.12 inches, when the larger box (1.40 inches) had slightly lower counts in the 2006 replication. With the increased range in size of the answer space tested in 2005 and 2006, there was no clear evidence of a ceiling effect and statistical analysis showed that the linear model provided a better fit than one with a polynomial included for box height (data not shown). It might be that still larger boxes would show diminishing returns for the length of answers. The results for the number of lines of text and number of sentences were nearly identical but are not shown here for space reasons.

Table 2. Effect of answer space size on an open-ended question, "Please explain why it did or didn't solve the problem or answer your question"

	Box height (inches)							
	.28	.56	.68	.84	1.12	1.40	1.68	
(*n*)	493	112	131	484	257	397	383	
Response probability								χ^2 *p*-value*
2003			49.6		45.8			.548
2004	49.6	58.9		48.1	46.0			.310
2005	51.4			50.0		53.9	48.2	.830
2006	47.2			45.5		47.5	48.5	.747
Pooled data	49.3	58.9	49.6	47.7	45.9	50.6	48.3	.651
(*n*)[a]	243	66	65	231	118	201	185	
Percent writing outside answer box								χ^2 *p*-value*
2003			13.9		3.0			.030
2004	12.7	7.6		7.8	1.9			.049
2005	14.9			7.8		3.8	4.4	.002
2006	21.3			2.2		5.2	6.4	.000
Pooled data	16.9	7.6	13.9	5.6	2.5	4.5	5.4	.000
Mean words								*p*-value†
2003			16.2		17.3			.074
2004	10.7	15.2		17.0	21.4			.000
2005	10.0			17.7		21.4	25.9	.000
2006	11.8			16.5		21.1	24.2	.000
Pooled data	10.9	15.2	16.2	17.1	19.1	21.3	25.0	.000

Table 2. Continued

	Box height (inches)							
	.28	.56	.68	.84	1.12	1.40	1.68	
(*n*)	493	112	131	484	257	397	383	
Mean themes								*p*-value ~
2003 data			1.28		1.27			.587
2004 data	1.24	1.23		1.29	1.62			.004
2005 data	1.15			1.21		1.25	1.29	.044
2006 data	1.15			1.29		1.31	1.35	.040
Pooled data	1.17	1.23	1.28	1.26	1.42	1.29	1.32	.009
Elaboration with additional details								χ^2 *p*-value*
2003 data			33.9		29.2			.706
2004 data	21.8	43.9		49.0	57.7			.000
2005 data	9.6			32.2		39.1	48.4	.000
2006 data	16.0			28.9		36.5	38.3	.000
Pooled data	14.8	43.9	33.9	34.6	41.9	37.8	43.2	.000

*Mantel-Haenszel Chi-square use for 2004–2006 and Pooled data. †*p*-value is based on generalized poisson regression. ~*p*-value is based on ordinal logistic regression.

[a] Percent writing outside answer box, means for words, lines of text, and sentences are calculated based on those providing a response.

Table 3. Effect of answer space size on an open-ended question, "What could we do to improve our service to you and others in the county?"

	Box height (inches)					
	.28	.56	.84	1.12	1.40	
(*n*)	495	647	243	494	378	
Response probability						χ^2 *p*-value*
2003		45.1	58.8			.030
2004	57.5	57.6	54.5	56.8		.798
2005	57.4	58.0		55.0	61.7	.638
2006	54.8	49.0		52.1	49.0	.449
Pooled data	56.4	52.2	56.8	54.3	55.0	.978
(*n*)[a]	279	338	138	268	208	
Percent writing outside answer box						χ^2 *p*-value*
2003		7.7	10.4			.772
2004	15.4	8.2	1.6	3.2		.003
2005	20.0	6.2		6.7	7.2	.008
2006	22.9	4.0		5.0	3.1	.000
Pooled data	20.7	6.2	6.5	5.2	5.3	.000
Mean words						*p*-value†
2003		10.4	14.1			.045
2004	9.3	13.5	14.2	19.2		.000
2005	8.9	12.6		18.1	19.1	.000
2006	9.7	10.7		20.4	16.6	.000
Pooled data	9.3	11.8	14.2	19.2	17.9	.000
Mean themes						*p*-value ~
2003 data		1.54	1.55			.991
2004 data	1.46	1.43	1.51	1.52		.647
2005 data	1.45	1.42		1.56	1.64	.030
2006 data	1.33	1.38		1.62	1.37	.240
Pooled data	1.40	1.43	1.53	1.58	1.52	.026
Elaboration with additional details						χ^2 *p*-value*
2003 data		15.4	40.3			.001
2004 data	16.9	34.4	31.2	39.7		.011
2005 data	13.3	29.2		35.6	46.0	.000
2006 data	14.7	24.2		30.7	28.9	.008
Pooled data	14.7	26.0	36.2	34.7	38.0	.000

*Mantel-Haenszel Chi-square use for 2004–2006 and Pooled data. †*p*-value is based on generalized poisson regression. ~*p*-value is based on ordinal logistic regression.

[a] Percent writing outside answer box, means for words, lines of text, and sentences are calculated based on those providing a response.

Next, the effect of the answer space size on content was examined. The number of themes generally increased with answer box height, as shown in the mean themes panels in Tables 2 and 3. As with the number of words, the initial comparisons in 2003 for two levels of box height had essentially no difference in the number of themes. The replications in

2004, 2005, and 2006 showed significant effects and the pooled data did as well for the explanation item. The pattern for the number of themes closely followed that for words in each of these years and for the pooled data. The results for the item about improving services were less clear-cut, with only the 2005 replicate and the pooled data revealing a significant relationship between box size and the number of themes. In addition, the analysis of the content category showed no clear relationship with box height for individual replications or the pooled data (data are not shown but are available, upon request, from the author).

Though the general content category of response did not appear to be affected by answer space size, there was evidence that the propensity to provide additional details in an answer was substantially impacted. As shown in the bottom panels of Tables 2 and 3, the percentage of respondents who provided additional details increased as the box grew taller for each replication except the 2003 version of the item "Please explain why it did or didn't solve the problem or answer your question." This finding is consistent with the results for the length of responses, which was reported earlier. It is noteworthy that there also was considerable variation between years (for example, 49.0% in 2004 and 28.9% in 2006 provided additional details for the .84 box height in Table 2).

Additional analysis using the pooled data only was conducted to examine whether the effect of box height was moderated by selected demographic attributes: education, age, and sex. As shown by the results in Table 4, age had a significant interaction with box size for the number of words and elaborating with additional details for the item "Please explain why it did or didn't solve the problem or answer your question." Using the parameter estimates to calculate predicted number of words, respondents over 60 years of age were more sensitive to box height than younger respondents, with the former writing the same number of words in the .28 inch high box, 2.2 words more in the .84 inch box, and 8.6 words more in the 1.68 inch box. Similarly, the percentage of older respondents who provided additional details was slightly lower than for younger respondents for the .28 inch box (22.7 and 25.0%, respectively) but considerably higher for the larger boxes (e.g., 64.5 and 40.4%, respectively for the 1.68 inch box). Finally, no interaction effects for box height and either education or sex were found (data not shown).

Though moderating effects of demographic attributes on the effect of box size were limited, there were important independent effects on answer quality. Sex had consistent significant effects for both items, in that women write more words and themes than men and are more likely to elaborate details and write outside the designated answer space. Education showed more limited effects, with college-educated respondents writing significantly more words and being more likely to write outside the designated answer space than those with less education on the improvement item.

Finally, the category of the primary theme of the response significantly influenced the number of words for both items, whether respondents elaborated for the improvement item, and whether respondents wrote outside the designated answer space for the explanation item. Regarding the explanation item, respondents who made a positive comment or provided a "don't know" type of response wrote significantly fewer words and were less likely to write outside the answer space than those writing a negative or other type of comment. Conversely, respondents with constructive suggestions for ways to

improve service wrote more, while those with negative comments and "don't know" wrote much less, relative to those with other comments on the improvement items. Respondents who elaborated on ways to improve services were more likely to make a positive, negative, or "don't know" type of response than to make an "other" comment.

Table 4. Parameter estimates for effects of answer space size and demographic attributes on selected outcomes for the pooled data, 2003–2006

	Outcome				
Please explain why it did or didn't solve the problem or answer your question	Response probability	Number of words in response	Number of themes	Elaborated with additional details	Writing outside the answer box
			− 1.523		
			− 3.599		
Intercept(s)[a]	− .060	2.609	− 6.583	− 1.102	− 1.296
Box size	− .035	**.450**	**.391**	**.517**	**− 1.206**
Sex (Male = 1)	**− .324**	**− .233**	**− .392**	**− .594**	**− .559**
Age (over 60 = 1)	.147	− .059	.152	− .348	**.654**
Education (college = 1)	**.380**	.060	.011	.214	.017
Box size x Age	–	**.221**	–	**.778**	–
Primary theme:					
Positive		**− .265**		− .211	**− .577**
Negative		**− .007**		.278	**.973**
Don't know response		**− .207**		.008	**− .656**
Other comment		**.000**		.000	**.000**
Dispersion parameter (a)		**.101**			
Model χ^2	29.875	218.3	14.353	75.790	62.727
Model Deviance	2,987.389	7,762.7	1,349.123	1,283.804	525.827
R^2_{Dev}	.010	.193	*n*/a	.056	.107
What could we do to improve our service to you and others in the county?					
			− .472		
			− 3.082		
Intercept(s)	.051	2.111	− 6.201	− 1.957	− 1.405
Box size	.022	**.604**	**.316**	**.981**	**− 1.536**
Sex (Male = 1)	− .028	**− .302**	**− .286**	**− .580**	**− .459**
Age (over 60 = 1)	.141	.025	**− .254**	− .045	− .222
Education (college = 1)	**.319**	**.116**	.116	.144	**.439**
Primary theme:					
Positive		**.210**		**.674**	.319
Negative		**− .674**		**.311**	− .821
Don't know response		**− .226**		**.255**	.215
Other comment		**.000**		**.000**	.000
Dispersion parameter (a)		**.144**			
Model χ^2	14.168	219.7	16.368	87.750	46.782
Model Deviance	2,974.326	8,466.4	2,049.995	1,367.613	691.388
R^2_{Dev}	.005	.161	*n*/a	.060	.063

[a] Parameter estimates for number of words are from generalized poisson regression; those for the number of themes are from ordinal logistic regression; those for response probability, elaboration with additional details and writing outside the answer box are from logistic regression. Bold type denotes p-value $< .05$. All models are significant with regard to intercept only models.

5. Discussion

The results from the experiments provide additional evidence that visual design elements, specifically the size of the answer space for open-ended questions, can have a substantial influence on respondents' behavior with regard to the answers to self-administered surveys. The results of the graphical manipulation of the answer space are consistent with earlier findings of Christian and Dillman (2004) and Stern et al. (2007), that the taller space encouraged respondents to write longer answers than a shorter space. The experiments here used an array of box sizes and showed that the length of the response, in terms of the number of words, increased as the vertical size of the answer space increased. Moreover, the increased size did not reduce the propensity to respond to the item. No evidence was found for a ceiling effect for the range of answer box sizes employed in the experiments.

In addition, larger answer spaces showed a tendency to elicit more themes and additional details. No evidence was found, however, that larger answer spaces generated more responses with a negative content in conjunction with longer answers, as might be predicted by Schwarz et al.'s (2003) work. The findings also are consistent with Christian and Dillman's (2004) study which showed the larger of two answer spaces generated more themes than the smaller one. On the other hand, the effect of answer space size for web-based surveys appears to depend on the motivation of respondents, with a more pronounced effect for late responders than for early responders (Smyth et al. 2009). Further testing of differences between mail and web surveys is needed.

From a practical standpoint, one might ask, "How big should the answer space be?" Selecting the appropriate size answer space for open-ended questions is important because a box that is too large might increase the perception that a lot of information is required (and hence information accessibility is difficult) while one that is too small might convey that a short, less detailed answer is expected. The latter is likely to generate responses with the most accessible, but less complete information that is less useful to the researcher (Schwarz et al. 2003). Though the expected length of an answer is likely to vary with the topic of the questions, in the case of the two questions included in this study, the largest answer space (which provided space for four-five lines of text) accommodated all but the most verbose respondents (2–3%) in the two experiments. Thus, the largest answer space was probably very close to the size for encouraging optimal responses – that is, responses that included additional details as well as the feelings of respondents.

While previous studies (Christian and Dillman 2004; Smyth et al. 2009) used samples of college students or residents in a single community (Stern et al. 2007), this study provides evidence to suggest that the visual impact of the answer space applies more broadly to the general population. As noted earlier, the characteristics of FCES clients are distributed somewhat differently than those of Florida's population. On the other hand, in the case of FCES clients there is included the diversity of age groups and education levels to test how these attributes might affect the box size – answer quality relationship in the general population. It is also important to note that the effects of answer space size were generally consistent across education, age, and gender categories. Though age moderated the influence of box size for one item, the negative effect on words and elaboration for small boxes was small for older persons relative to the gain predicted for the larger box sizes. Given a recommendation to use a larger answer space, the age moderator might cause

some concern about response quality in the case of younger persons. Further experimentation is needed to see how widespread the moderator effect might be. The lack of a pattern of demographic moderators suggests, however, that larger answer spaces should result in improved response quality (at least as measured by words, themes, and elaboration).

The results of this study, along with those of Christian and Dillman (2004) and Smyth et al. (2009), suggest that open-ended questions are feasible, at least with regard to questions that solicit relatively short responses. Researchers should also note that the range of topics that have been explored to date (student experiences, community change, and customer satisfaction) are limited. Whether these results hold for open-ended questions that inquire about other topics or that require paragraph-length or longer responses remains to be tested. Nevertheless, these studies can be used to guide specific decisions in developing survey instruments in order to maximize the completeness of answers. Surveys that incorporate the lessons from research on visual design effects should provide more useful information to policy makers, organizational leaders, and the public at large.

6. References

American Association for Public Opinion Research (2004). Standard Definitions Final Dispositions of Case Codes and Outcome Rates for Surveys. Accessed on June 25, 2005, at http://www.aapor.org/pdfs/standarddefs_ver3.pdf

Bishop, G.F., Hippler, H.-J., Schwarz, N., and Strack, F. (1988). A Comparison of Response Effects in Self-administered and Telephone Surveys. In Telephone Survey Methodology, R.M. Groves, P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls, and J. Waksberg (eds). New York: John Wiley and Sons, 321–340.

Cameron, A.C. and Windmeijer, F.A. (1996). R-square Measures for Count Data Regression Models with Applications to Health-Care Utilization. Journal of Business & Economic Statistics, 14, 209–220.

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

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

de Leeuw, Edith D. (2008). Choosing the Method of Data Collection. In International Handbook of Survey Methodology, E.D. de Leeuw, J.J. Hox, and D.A. Dillman (eds). NY: Lawrence Erlbaum Associates, 113–135.

Dillman, D.A. (2007). Mail and Internet Surveys: The Tailored Design Method, (2nd Edition). New York: John Wiley.

Dillman, D.A., Gertseva, A., and Mahon-Haft, T. (2005). Achieving Usability in Establishment Surveys Through the Application of Visual Design Principles. Journal of Official Statistics, 21, 183–214.

Gardner, W., Mulvey, E.P., and Shaw, E.C. (1995). Regression Analyses of Counts and Rates: Poisson, Overdispersed Poisson, and Negative Binomial Models. Psychological Bulletin, 118, 392–404.

Krosnick, J.A. (1999). Survey Research. Annual Review of Psychology, 50, 537–567.

Krosnick, J.A., Narayan, S., and Smith, W.R. (1996). Satisficing in Surveys: Initial Evidence. New Directions for Evaluation, 70, 29–44.

Long, J.S. (1997). Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks, CA: Sage Publications.

Mittlböck, M. and Schemper, M. (1996). Explained Variation for Logistic Regression. Statistics in Medicine, 15, 1987–1997.

Mittlböck, M. and Schemper, M. (1999). Computing Measures of Explained Variation for Logistic Regression. Computer Methods and Programs in Biomedicine, 58, 17–24.

Mittlböck, M. and Waldhör, T. (2000). Adjustments for R^2-measures for Poisson Regression Models. Computational Statistics & Data Analysis, 34, 461–472.

Redline, C., Dillman, D.A., Dajani, A.N., and Scaggs, M.A. (2003). Improving Navigational Performance in U.S. Census 2000 by Altering the Visually Administered Languages of Branching Instructions. Journal of Official Statistics, 19, 403–419.

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

Schuman, H. and Presser, S. (1981). Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context. Orlando, FL: Academic Press.

Schwarz, N., Bless, H., Wänke, M., and Winkielman, P. (2003). Accessibility Revisited. Foundations of Social Cognition: A Festschrift in Honor of Robert S. Wyer, Jr, GalenV. Bodenhausen and Alan J. Lambert (eds). Mahwah, NJ: L. Erlbaum, 51–77.

Smyth, J.D., Dillman, D.A., Christian, L.M., and Stern, M.J. (2004). How Visual Grouping Influences Answers to Internet Surveys. Revision of a paper presented at the 2004 Annual Meeting of the American Association for Public Opinion Research, Phoenix, AZ, May 13.

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

Stern, M.J., Dillman, D.A., and Smyth, J. (2007). Visual Design, Order Effects, and Respondent Characteristics in a Self-Administered Survey. Survey Research Methods, 1, 1–11.

Strauss, A.L. and Corbin, J.M. (1990). Basics of Qualitative Research: Grounded Theory and Techniques. Newbury Park, CA: Sage Publications.

Tourangeau, R. (2004). Survey Research and Societal Change. Annual Review of Psychology, 55, 775–801.

Tourangeau, R. and Rasinski, K.A. (1988). Cognitive Processes Underlying Context Effects in Attitude Measurement. Psychological Bulletin, 103, 299–314.

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

Wang, W. and Famoye, F. (1997). Modeling Household Fertility Decisions with Genralized Poisson Regression. Journal of Population Economics, 10, 273–283.

Received June 2008
Revised October 2009