THE EFFECTS OF ALTERING THE DESIGN OF BRANCHING INSTRUCTIONS ON NAVIGATIONAL PERFORMANCE IN CENSUS 2000

Cleo Redline, U.S. Bureau of the Census, Don Dillman, Washington State University, Aref Dajani, and Mary Ann Scaggs, U.S. Bureau of the Census
Cleo Redline, U.S. Bureau of the Census, Room 3215—4, Washington, DC 20233

Keywords: Questionnaire Design; Measurement Error; Errors of Commission and Omission; Branching Instructions

It has been hypothesized that a number of languages (verbal, symbolic, and graphic paralanguage) combine to affect respondent perception and comprehension of branching instructions, and consequently, the navigational path respondents follow when completing a questionnaire. A pilot study with college students in which these languages were altered in two distinct ways (the *prevention* and *detection* methods) and tested against the Census 2000 method of branching provided evidence for this proposition (Redline and Dillman In Press).

This paper reports on an experiment conducted in Census 2000 in which the two branching instructions from the classroom experiment were revised, and two additional instructions were developed (*reverse printing* the instruction and substituting the words "*go to*" for "*skip to*") and tested against the Census 2000 version. In this paper, we report whether altering the languages of the branching instructions in the Census 2000 long form had an effect on mail response rates or navigational errors. This experiment was designed to extend the findings of the classroom experiment and to provide insight about the effects of altering the design of branching instructions with a diverse population and when the wording of the questions themselves can provide navigational clues, in addition to the branching instructions.

Background

Information on a self-administered questionnaire can be decomposed into four language types: verbal, numeric, symbolic, and graphic paralanguage (Redline and Dillman In Press).

- Verbal—refers to the words
- Numeric—refers to the numbers
- Symbolic—refers to the check boxes, arrows and other symbols on the questionnaire
- Graphic paralanguage—is the conduit by which all of the other languages are conveyed and includes the brightness, color, shape, and location of the information

The major thesis of this program of research is that these languages combine to create meaning for respondents, and that with conventional branching instructions, three of these languages (the verbal, symbolic, and graphic paralanguage) combine in such a way that respondents are often left unaware of the branching instruction.

One reason for this may be that, typically, these instructions are printed in the same font and point size as the rest of the text, making them difficult to detect (Wallschlaegar and Busic-Snyder 1992). In addition, Kahneman (1973) demonstrated that people's vision is sharp only within 2 degrees, which is equal to about 9 characters of text. Consequently, when a respondent is in the process of marking a check box, the branching instruction, which is usually located to the right of a response option, is likely to be outside of the respondent's view. Also, this design does not take into consideration other strategies for reducing human error, like training respondents to prevent their errors in advance, or allowing them to detect errors afterwards (Norman 1990; Wickens 1992).

Thus, two new designs, the prevention and detection branching instructions, were developed. Redline and Dillman (In Press) present a detailed description of these instructions, along with their depiction, which is briefly summarized here. In the prevention method, an instruction was placed before the question to remind respondents to pay attention to the branching instructions. The purpose of these reminders was to prevent mistakes before they were made. Also, the location of the response options and check boxes were reversed to bring the branching instruction into view and the branching instruction was made larger and bolder and it was placed in a white background. In the detection design, the branching instruction was made even larger and bolder to compensate for its conventional location. Also, a left-hand arrow came off of the non-branching response options and pointed to a parenthetical feedback phrase. The purpose of these phrases was to allow respondents to detect and correct their mistakes after they had made them. Consequently, both instructions attempted to make the verbal skip instruction more visible, but they differed in that the prevention technique tried to remind people in advance that they may need to branch, whereas the detection technique gave them information afterwards, which allowed them to determine if they had branched correctly.

In a classroom experiment of 1,266 students, these designs were shown to decrease errors of commission (respondents answering questions they were instructed to skip) by more than half. However, errors of omission (respondents not answering questions they were instructed to answer) increased (Redline and Dillman In Press). In addition to the experiment, 48 cognitive interviews were conducted with a broad mix of people (Dillman et al. 1999; Redline and Crowley 1999).

Although respondents were supposed to understand that the check box and branching instruction were connected because they were next to each other in the same white background in the prevention branching instruction, the pretests suggested that this did not work. Consequently, a stronger visual connection, an arrow, was placed between the two in the census experiment. The reminder instructions may have contributed to the increased errors as well. Therefore, the number of these was dramatically reduced by strategically placing them after a long series of questions without any branching instructions.

The larger size of the detection branching instruction in the classroom experiment appeared to overly attract respondents' attention to it, so this instruction was decreased in size for the census experiment. Also, respondents had trouble when they came to a branching instruction at the bottom of a page because the left-hand arrow did not point to anything. Thus, the left-hand arrow was made to terminate into another instruction at the bottom of a page in the census design.

Also, two new instructions were developed to test additional issues. A branching instruction was designed to test the hypothesis that simply changing the verbal language from "skip to" to "go to" without making the instruction more visible is unlikely to make a difference in respondents' performance.

Another branching instruction was designed to provide insight into whether printing branching instructions in reverse print is a good practice or not. Normal print is the black lettering on the yellow background typical of most information on the census questionnaires. Reverse print is yellow lettering on a black background. There are arguments both for and against using reverse print. On the one hand, it is plausible that the high contrast of a reverse-printed branching instruction and the fact that it is made visually dissimilar from the other information on the questionnaire could attract respondents' attention. On the other hand, typographical studies warn against using reverse print because it is difficult to read (Hartley, 1981; Wallschlaeger and Busic-Snyder, 1992).

Experimental Design and Implementation Procedures

These ideas were tested in an experiment in Census 2000 using the long form. Five versions of the long form were developed, each employing a different treatment of the branching instruction. Addresses on the Decennial Master Address File in the mailout/mailback areas of the country at the time sample selection took place served as the universe for sample selection (Woltman, 1999). A sample of approximately 25,000 addresses was selected to receive one of the five treatments, with approximately 5,000 addresses independently selected per treatment. This number was distributed equally between so-called high coverage areas (2,500 per treatment), which are expected to have a low proportion of minorities and renters, and low coverage areas (2,500 per treatment), which are expected to have a high proportion of minorities and renters. The five treatments were:

TREATMENTS

1. <u>The Census 2000 Skip To Instruction.</u> Shown in Figure 1a, this instruction was used in the classroom experiment, and is exactly the same as the instruction used on the Census 2000 long form.
2. <u>The Go To Instruction.</u> Shown in Figure 1b, this instruction is like the Census 2000 instruction in all respects, except that the words "skip to" have been changed to "go to."
3. <u>The (Go To) Reverse Print Instruction.</u> Shown in Figure 1c, this instruction is like the Go To instruction, except that the words 'Go to" have been changed from normal print (black lettering on a yellow background) to reverse print (yellow lettering on a black background).
4. <u>The (Go To) Prevention Instruction</u>. Shown in Figure 1d, "skip to" was changed to "go to." A bold arrow was placed between the check box and the branching instruction to make the connection between these two pieces of information clearer. Also, the number of reminder instructions was dramatically reduced.
5. <u>The (Go To) Detection Instruction</u>. Shown in Figure 1e, "skip to" was changed to "go to." The size of the branching instruction was decreased slightly from what it had been in the classroom experiment, and a left-hand arrow that terminated into a verbal branching instruction at the bottom of pages was added.

ANALYTIC PROCEDURES

<u>Calculating Mail Response Rates</u>

Households that returned duplicate forms were excluded from the calculation of the mail response rate (2 cases), as were households who did not return a form, but who were subsequently labeled as undeliverable as addressed in the mailout file (2834 cases). It was assumed in the latter case that the household was correctly labeled as non-existent or vacant. However, households that were identified as undeliverable as addressed in the mailout file, but who returned a questionnaire were included in the calculation. It was assumed in this case that the household was mistakenly labeled in the mailout file.

Nonresponse, then, was defined as any remaining household in the mailout universe who did not return a form, or who returned a blank form. Blank forms were defined as having less than two answers for the first two persons per form. Response was defined as households from whom a non-blank form from the mailout universe was received. The aggregate total for all responses and for all nonresponses was established and then the total number of responses was divided by the total number of responses plus the total number of nonresponses to yield the mail response rate.

Calculating Error Rates

To control for differences in the number of questions that respondents answered, this analysis was limited to the questions for Person 1. Branching error rates were calculated for questions that had branching instructions (because only their designs differed between treatments) and those questions that had valid responses (because only then was it evident whether a respondent should branch or not).

An opportunity to make an error of commission occurred when a respondent selected a response with a branching instruction associated with it. An error of commission occurred if the respondent answered a question other than the question they were instructed to answer. An opportunity to make an error of omission occurred when a respondent selected a response that did not have a branching instruction associated with it. An error of omission occurred if the following question was left blank. Commission and omission opportunities, errors, and rates were calculated by respondent, by question within a treatment, across all questions, and across all treatments.

Significance Testing

Households were sampled randomly at different rates within two geographic strata: high coverage areas and low coverage areas. Branching error rates were calculated by dividing the number of branching errors by the number of branching opportunities, where each of the two quantities is random. To compare rates across strata or treatments, standard errors were calculated using the statistical replication method of the stratified jackknife. As each household can have a variable number of branching opportunities and errors, clusters were incorporated into the variance estimation at the household level. Operationally, the stratified jackknife dropped one household at a time to calculate variance estimates. Statistical significance testing was conducted on pairs of strata or treatments using a t-test that incorporates the covariance between the branching opportunities and branching errors in the calculation. The normal approximation to the t distribution was used to calculate p-values and establish statistical significance. A Bonferroni adjustment was used to account for the multiple comparisons between treatments.

Results and Discussion

RESPONSE RATES

Response rates for the five treatment groups varied significantly between the high and low coverage areas, averaging 66.7% for the former and only 48.6% for the latter, for a difference of about 18 percentage points, which is significant at the .01 level of the test. We hypothesized that changes in the branching instructions would have little effect on the response rates *between treatment groups*, and for the most part this is true. One exception is the Reverse Print treatment, which is significantly different from the Go To treatment at the 0.05 level of the test. Perhaps respondents thought the Reverse Printed form looked more difficult to complete. However, the reason for this is not perfectly clear, so we are reserving our judgment concerning this finding until it is replicated.

When the data are parsed by high and low coverage area, the patterns are nearly identical, suggesting that for the most part, treatment and coverage area do not interact.

COMMISSION ERROR RATES

It can be seen in Table 1 that the average commission error rate for the 19 branching items contained in the Census 2000 Skip To form (Treatment 1) was high. It was 19.7%. As expected, the rate was substantially higher in the low coverage areas (26.9%) compared to the high (18.6%).

The commission error rate for the Go To comparison (Treatment 2) was not significantly different from the Skip To treatment for all areas. Thus, the results of this experiment provide support for our original hypothesis that changing the instruction from “Skip to” to “Go to” would not affect the error rate because such a change does not address the underlying need to attract respondents’ attention to the instruction first.

Because all of the remaining treatment groups used the words “Go to,” Treatment 2 becomes the control group for the remaining comparisons. The fact that the commission error rate decreases across the Go To (20.8%), Reverse Print (17.9%), Prevention (14.7%) and Detection (13.5%) treatments suggests that for the most part, the changes made from one design to the next increasingly improved respondents’ perception and comprehension of the instruction. Most of the error rates are significantly different from one another at the .01 level of the test, except the for the Prevention and Detection treatments in the high coverage area, which were not significantly different from one another.

Although the Reverse-Printed treatment led to a reduction in the commission error rates, this reduction was mediocre in comparison to the Prevention and Detection treatments. A plausible explanation for this is that respondents get used to reading information in a particular figure-ground (in this case, black against a yellow background). As a result, they come to expect that the information they should pay attention to will be black against yellow. When the instruction is reverse printed, it may look so different that respondents tend to disregard it. The Gestalt Grouping Law of Similarity states that we tend to see similar elements as belonging together (Wallschlaeger and Busic-Snyder, 1992), and here the visual elements of color are not similar. So this may be an example of not using the visual element of color, or the Gestalt Grouping Law of Similarity, to our benefit.

The improvement in performance between both the Prevention and Detection treatments over the other treatments is likely due to making the branching instruction more visible. However, the added improvement the Detection treatment displayed over the Prevention treatment in the low coverage area may be due to the feedback mechanism.

OMISSION ERROR RATES

The omission error rate averaged about 5 percent for the nation as a whole. Once again it was higher in the low coverage area (6.5%) than the high (4.8%), and as predicted there weren't any differences in the error rate between the skip to and go to instructions. The other important finding here is that only the Detection form significantly reduced the omission error rates (4.0%), whereas the Reverse Print and Prevention increased it (7.6% and 7.0%, respectively).

These results suggest that at the same time strong steps are taken to visually associate (or group) the check box(es) with the branching instruction(s), counter steps must also be taken to clearly disassociate the branching instruction from the other response options (i.e., not allow them to be seen as grouped together). It would seem that the Detection Treatment accomplished this balancing (grouping) act best.

In addition, the feedback mechanism may have worked better in the census experiment than the classroom experiment because it was simpler. In the census experiment it was almost always "(If Yes)" or "(If No)," whereas in the classroom experiment it tended to be a more complicated phrase, like "(If basketball, wrestling, or sent here from an earlier question).

EFFECTS OF CUES FROM FOLLOW-UP QUESTIONS

The classroom experiment controlled for the effects of the wording of the questions so that respondents could get no cues from the questions themselves whether they should be answering them. However, in the census the questions were dependent. So, for example, one of the questions asked respondents if they had any of their own grandchildren under the age of 18 living in their house or apartment, and if they did, then they were asked a follow-up question concerning whether they were responsible for these grandchildren. It seemed reasonable to expect that respondents would be able to figure out if a follow-up question applied to them in the census, *not* from reading the branching instruction, but from reading the content of the follow-up question and that the error rates would be lower in the census as a result.

However, the error rates are either the same or higher in the census compared to the classroom. This suggests that nationally representative respondents have a tendency to answer questions that do not apply to them, despite the fact that the questions themselves may provide them with contrary cues and contain branching instructions telling them to do otherwise. Therefore, not only may it be a good idea to improve upon the branching instruction, but it may pay to improve respondents' comprehension of the questions as well. Steps taken to limit navigational errors reduces respondent burden.

Conclusion

This paper provides the first clear evidence from a field experiment that using visual and human performance theory to design a questionnaire's branching instructions affects respondents' navigational performance. We have shown that simultaneously manipulating the graphic paralanguage, symbolic, and verbal languages that comprise branching instructions influences significantly whether those instructions are followed. Thus, our general hypothesis that more than one language affects how respondents navigate and complete questionnaires is confirmed. The results of this experiment suggest, however, that learning to manipulate these languages so that the information is grouped correctly is critical to successfully accomplishing this task.

The failure of nearly 20% of respondents, on average, to follow branching instructions in Census 2000 strikes us as unacceptably high. The Detection Treatment reduced commission errors by about one-third and omission errors by about one-fourth. This method of providing branching instructions seems usable for most self-administered questionnaires in its present form, though further improvements may be possible.

It is also apparent that wide variations existed in the error rates for individual questions. Thus, the potential reasons for these variations will be analyzed in future papers, as will the relationship between respondent characteristics and branching errors. Left for future research also is the potential we see for combining the

Prevention and Detection Treatments in order to harness the theoretical power that certain aspects of each seems to offer for improving performance with branching instructions. Finally, the major thesis of research--that the graphic paralanguage, symbolic, numeric, and verbal languages of a questionnaire combine to create meaning for respondents—now needs to be systematically extended to (that is, tested on) other areas of the questionnaire, the most interesting of which will undoubtedly be the questions themselves.

Acknowledgement

This paper reports the results of research and analysis undertaken by Census Bureau and Washington State University staff. It has undergone a review more limited in scope than that given to official Census Bureau publications. This paper is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

References

Dillman, D., Carley-Baxter, L., and Jackson, A. 1999. Skip Pattern Compliance in Three Test Forms: A Theoretical and Empirical Evaluation. SESRC Technical Report #99-01 Social and Economic Sciences Research Center. Pullman: Washington State University.

Jenkins, C. and Dillman, D. 1997. "Towards A Theory of Self-Administered Questionnaire Design." In L. Lyberg, P. Biemer, M. Collins, L. Decker, E. DeLeeuw, C. Dippo, N. Schwarz and D.Trewin (eds), *Survey Measurement and Process Quality*. NewYork: Wiley-Interscience.

Hartley, J.1981. Eighty ways of Improving Instructional Text. IEEE Transactions of Professional Communication, 24, 1, 17-27.

Kahneman, D. 1973. *Attention and Effort*. New Jersey: Prentice Hall.

Norman, D. 1992. *The Design of Everyday Things*. New York: Currency Doubleday.

Redline, C. and Dillman, D. In Press. "The Influence of Alternative Visual Designs of Respondents' Performance with Branching Instructions in Self-Administered Questionnaires" In Groves, R., Dillman, D., Eltinge, E., and Little, R. (eds.) *Survey Nonresponse*. New York: John Wiley and Sons, Inc.

Redline, C., Dillman, D., Smiley, R., Carley-Baxter, L. Jackson, A. 1999. Making Visible the Invisible: An Experiment With Skip Instructions on Paper Questionnaires. Proceedings of the Section on Survey Methods Research, American Statistical Association.

Redline, C. and Crowley, M. 1999. Unpublished Data. Washington, D.C.: U.S. Bureau of the Census.

Wallschlaeger, C. and Busic-Snyder, C. 1992. *Basic Visual Concepts and Principles.* Dubugue, Iowa: Wm.C.Brown Publishers.

Wickens, C. 1992. *Engineering Psychology and Human Performance.. Second Edition.* HarpersCollins Publishers, Inc.

Woltman, H. 1999. Sampling Specifications for the Research and Experimentation Program in K. Shaw, *Program Master Plan for the Census 2000 Alternative Questionnaire Experiment (AQE2000).* U. S. Bureau of the Census, Planning, and Research Evaluation Division, December 22, 1999.

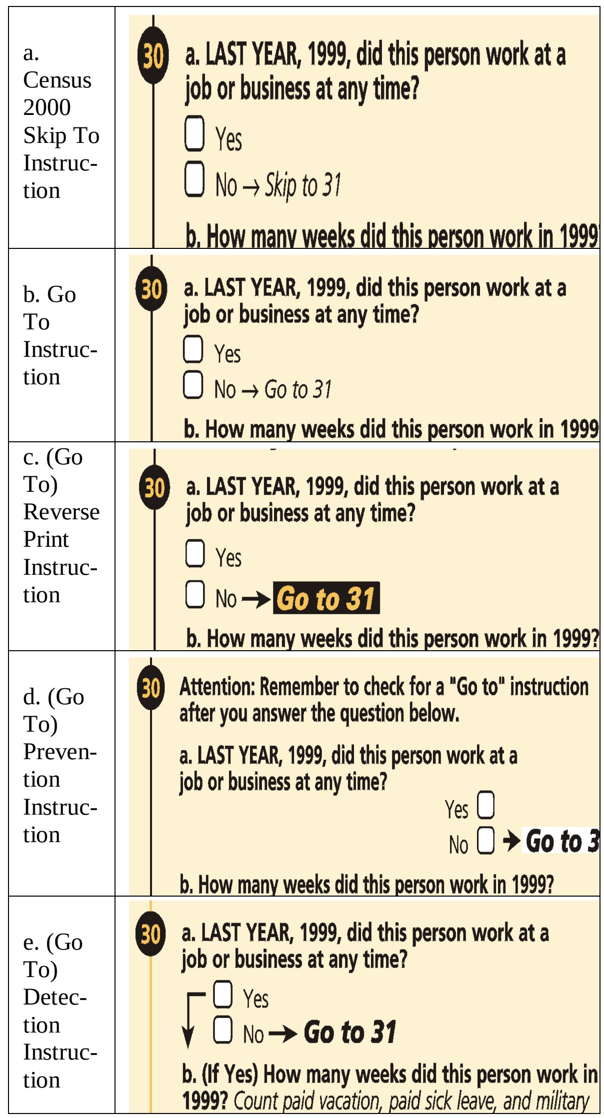

Figure 1. Illustration of the five branching instruction treatments.

Table 1. Commission Error Rates For All Census Long-Form Items With Branching Instructions

Instruction Treatment	Weighted National Total	High Coverage Areas	Low Coverage Areas
1. Census 2000 Skip To	19.7%	18.6%	26.9%
2. Go To Control	20.8	20.0	25.4
3. (Go To) Reverse Print	17.9	16.7	24.9
4. (Go To) Prevention	14.7	13.6	21.7
5. (Go To) Detection	13.5	12.7	18.6
Statistical Comparison			
1 vs. 2	n.s.	n.s.	n.s.
2 vs. 3	p < .01	p < .01	n.s.
2 vs. 4	p < .01	p < .01	p < .01
2 vs. 5	p < .01	p < .01	p < .01
3 vs. 4	p < .01	p < .01	p < .01
3 vs. 5	p < .01	p < .01	p < .01
4 vs. 5	n.s.	n.s.	p < .01

Table 2. Omission Error Rates For All Census Long-Form Items With Branching Instructions

Instruction Treatment	Weighted National Total	High Coverage Areas	Low Coverage Areas
1. Census 2000 Skip To	5.0%	4.8%	6.5%
2. Go To Control	5.4	5.2	6.3
3. (Go To) Reverse Print	7.6	7.3	9.1
4. (Go To) Prevention	7.0	6.7	9.4
5. (Go To) Detection	4.0	3.7	6.2
Statistical Comparison			
1 vs. 2	n.s.	n.s.	n.s.
2 vs. 3	p < .01	p < .01	p < .01
2 vs. 4	p < .01	p < .01	p < .01
2 vs. 5	p < .01	p < .01	n.s.
3 vs. 4	n.s.	n.s.	n.s.
3 vs. 5	p < .01	p < .01	p < .01
4 vs. 5	p < .01	p < .01	p < .01