Journal of Official Statistics, Vol. 19, No. 4, 2003, pp. 403–419

Improving Navigational Performance in U.S. Census 2000 by Altering the Visually Administered Languages of Branching Instructions

Cleo Redline[1], *Don A. Dillman*[2], *Aref N. Dajani*[3], *and Mary Ann Scaggs*[4]

This article reports results from an experiment in which the branching instructions in the long form of the 2000 U.S. Decennial Census were varied in a national experiment administered to 25,000 households. The experimental manipulations included testing a simple change in the verbal language of the current design from ‘‘Skip to’’ to ‘‘Go to.’’ Three additional combined manipulations were also tested in which the graphic (e.g., larger bolder font), symbolic (e.g., arrows), and verbal (e.g., additional statements) were altered. Whereas the simple change in the verbal language only (from skip to go) had no effect on branching error rates, the combined manipulations of the visually administered languages tended to reduce the branching errors. Results of these experiments provide evidence that errors in following branching instructions can be reduced for general populations through the careful manipulation of the visually administered languages.

Key words: Questionnaire design; errors of commission; errors of omission.

1. Introduction

In a recent classroom experiment with college students, it was found that different combinations of visually administered languages (verbal, symbolic, and graphic) affected respondents' perceptions and comprehension of branching instructions. That experiment revealed that two strategies for manipulating the visually administered languages of branching instructions, labeled the detection and prevention methods, were both effective in reducing the number of commission errors (failing to skip ahead when directed), but not the number of omission errors (skipping ahead when directed not to do so) (Redline and Dillman 2002). However, that experiment was limited by its being restricted to college students who completed the test questionnaires in a classroom setting. In addition, because one of the goals of the classroom experiment was to provide a pure test of the visual

[1] National Science Foundation, Division of Science Resources Statistics, 4201 Wilson Blvd., Suite 965, Arlington, VA 22205, U.S.A. Email: credline@nsf.gov
[2] Social and Economic Sciences Research Center and Departments of Sociology and Rural Sociology, Washington State University, Pullman, WA 99164-4014, U.S.A. Email: dillman@wsu.edu
[3] U.S. Bureau of the Census, Statistical Research Division, 4700 Silver Hill Road Stop 9100, Washington, D.C. 20233-9100, U.S.A. Email: aref.n.dajani@census.gov
[4] U.S. Bureau of the Census, Statistical Research Division, 4700 Silver Hill Road Stop 9100, Washington, D.C. 20233-9100, U.S.A. Email: mary.ann.scaggs@census.gov

Acknowledgments: This article reports research and analysis undertaken by U.S. Census Bureau and Washington State University staff. It has undergone a more limited review than official U.S. Census Bureau publications. It is released to inform interested parties of research and to encourage discussion. We would like to thank Elizabeth Martin, David Raglin, and Kristin Stettler, along with the Associate Editor, and two anonymous reviewers for their constructive comments on earlier drafts of the article.

language effects, none of the questions provided subject matter clues as to whether the next listed question should or should not be answered. Thus, although the percentage of commission errors was fairly large (20.3 percent for the control and 9.0 and 7.4 percent for the experimental forms), it was unclear whether the results would be similar in a typical questionnaire in which the questions often provide clues as to whether they should be answered (e.g., ''Are you employed?'' followed by ''How many hours do you work each week?'').

Our purpose in this article is to report results from an expanded test of these visual language concepts in which a national sample of U.S. households completed questionnaires delivered to their homes. In addition, the effects of two previously untested changes, one verbal (changing instructions ''skip to'' to ''go to'') and one graphic (reverse printing of the branching instructions), were also evaluated.

In a larger sense, this research addresses an oft-reported problem in the completion of self-administered questionnaires – the fact that questionnaires that include branching instructions are more likely to have higher item nonresponse rates to questions that should be answered (Featherston and Moy 1990).

2. Background

A data collection instrument that a respondent self-completes through the visual channel, such as on paper or over the Web, is visually administered (Redline and Lankford 2001). Visually administered questionnaires comprised both verbal and nonverbal languages, with the nonverbal comprising numeric, symbolic, and graphic languages (Redline and Dillman 2002).

- Verbal language – refers to the words.
- Numeric – refers to the numbers.
- Symbolic – refers to the check boxes, arrows, and other symbols.
- Graphic – is the visual 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 research is that respondents make mistakes responding to visually administered questionnaires not only because they do not understand the verbal language of a questionnaire, but because they do not understand the nonverbal (numeric, symbolic, and graphic) languages as well. Branching instructions were chosen as a departing point for studying this proposition because branching instructions yield a measurable truth (i.e., a respondent either follows the branching instruction correctly or does not) and because there is a mounting body of literature linking branching instructions with respondent errors (e.g., Messmer and Seymour 1982; Featherston and Moy 1990; Turner et al. 1992).

One reason respondents may make mistakes with branching instructions is that skip patterns are a concept from interviewer-administered protocols with which self-administered respondents may not be familiar. As shown in Figure 1a, typically these instructions are printed in the same font and point size as the rest of the text, making them difficult to detect (Foster 1979). 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

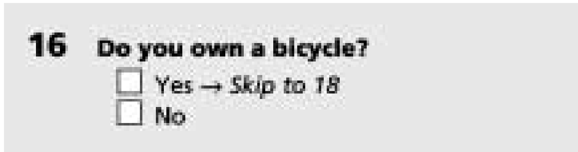
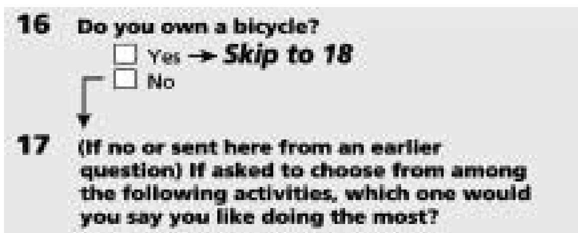

16 Do you own a bicycle?
☐ Yes → Skip to 18
☐ No

Fig. 1a. The Skip To instruction from the classroom experiment


16 Attention: Check for a skip after you answer . . .
Do you own a bicycle?
Yes ☐ Skip to 18
No ☐

Fig. 1b. The Prevention instruction from the classroom experiment


16 Do you own a bicycle?
☐ Yes → Skip to 18
☐ No

17 (If no or sent here from an earlier question) If asked to choose from among the following activities, which one would you say you like doing the most?

Fig. 1c. The Detection instruction from the classroom experiment

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, which manipulated the brightness, color, shape, and location (the visual design) of the branching instructions, as well as incorporating prevention and detection strategies. A detailed description of the development of these instructions is offered in Redline and Dillman (2002) and briefly summarized here. In the prevention method, an instruction was placed before each question with a branching instruction to remind respondents to pay attention to the latter instruction. On the first page, the instruction read ''Attention: Check for a skip instruction after you answer the question below'' but was abbreviated to the phrase shown in Figure 1b on subsequent pages. The purpose of these reminders was to prevent mistakes. Also, the locations 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.

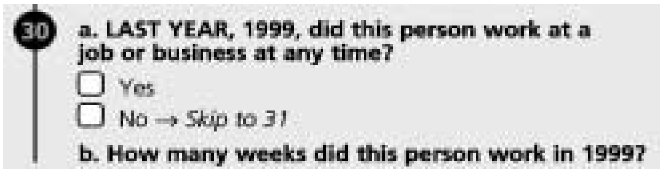
In the detection design, the branching instruction was made even bolder and larger to compensate for its poor location (see Figure 1c). Also, a left-hand arrow came off of the nonbranching response options and pointed to a parenthetical phrase. The purpose of these phrases was to allow respondents to detect and correct their mistakes. Consequently, both branching instruction methods 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 might need to branch, whereas the detection technique gave them information afterwards, which allowed them to determine if they had branched correctly.

In a previously reported classroom experiment involving 1,266 students, both designs were shown to decrease errors of commission (respondents answering questions they were instructed to skip) by more than half, from 20.3 percent on the control to 9.0 percent on the prevention and 7.4 percent on the detection form. However, errors of omission (respondents not answering questions they were instructed to answer) increased from 1.6 percent on the control to 3.3 percent on the prevention and 3.7 percent on the detection form (Redline and Dillman 2002).

In addition to the classroom experiment, 48 cognitive interviews were conducted with a broad mix of people, which provided insight into the potential reasons for branching errors (Dillman et al. 1999; Redline and Crowley 1999), and which led to both the development of new instructions and the refinement of existing ones for testing in the census. For example, respondents in the cognitive interviews suggested that ''go to'' was clearer than ''skip to.'' Thus, a branching instruction was designed in which the words ''go to'' replaced the words ''skip to'' (see Figure 2b). However, it was hypothesized that simply changing the verbal language from ''skip to'' to ''go to'' without making the instruction more visible was unlikely to make a difference in respondents' performance. Thus, the purpose of the ''go to'' instruction was to demonstrate that rewording information might not prove to be the answer if the underlying problem is that respondents are not reading information in the first place. This is not to say, however, that rewording will not prove useful once respondents read the information. Rewording may very well improve comprehension, and consequently performance, once respondents read the information, which explains the use of the words ''go to'' in all the experimental versions of the instruction.

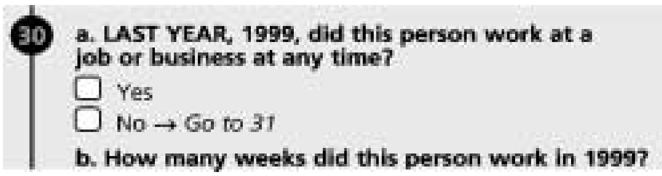
In addition to the ''go to'' instruction, another instruction was developed to provide insight into whether printing branching instructions in reverse print is a good practice or not (Figure 2c). Normal print is the black lettering on the light orange background typical of most information on the census questionnaires, or a light gray background when presented in black and white print, as is the case in Figure 2. Reverse print is orange lettering (which, again, is depicted as gray in Figure 2) on a black background, that is, a figure and ground change. There are arguments both for and against using reverse printing. 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 (Foster 1979). On the other hand, typographical studies warn against using reverse print because it is difficult to read (Hartley 1981; Wallschlaeger and Busic-Snyder 1992). Also, since most of what respondents generally read is black, they may come to expect information to be printed in black. As a result, they may pay less attention to the occasional reverse-printed instruction.

Although respondents were supposed to understand that the check box and branching instructions were connected because they were next to each other in the same white background in the prevention instruction, both the results of the classroom experiment and the cognitive interviews suggested that this did not reliably work. Consequently, a stronger visual connection, an arrow, was devised for the census experiment (see Figure 2d).



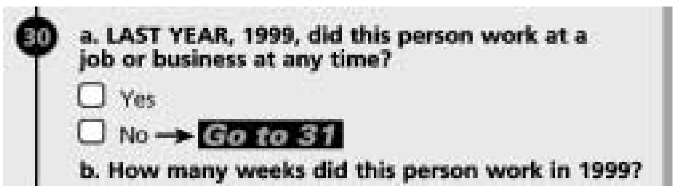
30 a. LAST YEAR, 1999, did this person work at a job or business at any time?
☐ Yes
☐ No → *Skip to 31*
b. How many weeks did this person work in 1999?

Fig. 2a. The Skip To instruction from Census 2000



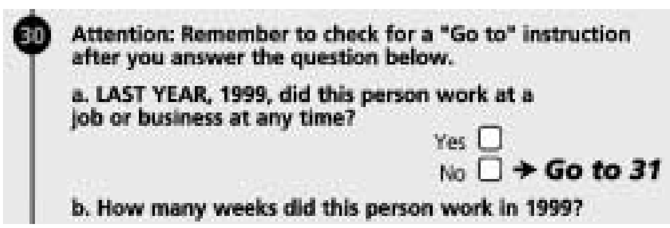
30 a. LAST YEAR, 1999, did this person work at a job or business at any time?
☐ Yes
☐ No → *Go to 31*
b. How many weeks did this person work in 1999?

Fig. 2b. The Go To instruction from Census 2000



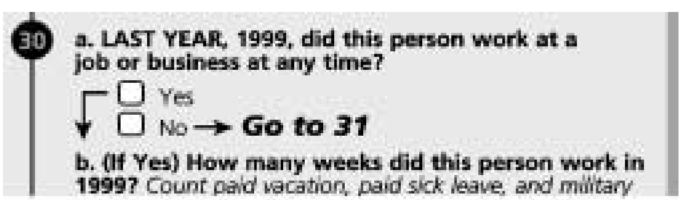
30 a. LAST YEAR, 1999, did this person work at a job or business at any time?
☐ Yes
☐ No → **Go to 31**
b. How many weeks did this person work in 1999?

Fig. 2c. The Reverse Print instruction from Census 2000



30 Attention: Remember to check for a "Go to" instruction after you answer the question below.
a. LAST YEAR, 1999, did this person work at a job or business at any time?
Yes ☐
No ☐ → ***Go to 31***
b. How many weeks did this person work in 1999?

Fig. 2d. The Prevention instruction from Census 2000



30 a. LAST YEAR, 1999, did this person work at a job or business at any time?
☐ Yes
☐ No → ***Go to 31***
b. (If Yes) How many weeks did this person work in 1999? *Count paid vacation, paid sick leave, and military*

Fig. 2e. The Detection instruction from Census 2000

Also, the cognitive interviews suggested that the number of reminder instructions might have contributed to the increased omission error rate for the prevention method that was evident in the classroom experiment. Therefore, the number of reminder instructions was dramatically reduced by selectively placing a reminder instruction after a long series of questions without any branching instructions for the census experiment.

Finally, 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 (see Figure 2e). 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 verbal branching instruction at the bottom of a page in the census design.

3. Methodology

These ideas were tested in a nationwide experiment embedded in the 2000 Decennial Census of the United States (Census 2000). Five versions of the long form were developed, each employing a different treatment of the branching instruction. The long form contains a total of 38 pages and asks 92 questions of each person living or staying at an address for up to six people per address. Of the 92 questions per person, 19 contained branching instructions.

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 high response rate, and low coverage areas (2,500 per treatment), which are expected to have a low response rate.

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. Consequently, addresses in nonmailback areas of the country (which can be characterized as highly rural areas, where the forms need to be dropped off and picked back up by interviewers) were excluded from the sample. Also, addresses that were added later as a result of coverage improvement operations were not included because these addresses were unavailable at the time of sample selection. Furthermore, addresses that were in the accuracy and coverage evaluation were excluded from the sample so as not to overburden these households. A systematic sample by state, stratum (the high coverage and low coverage areas), and treatment was selected (Woltman 1999). Analysis, then, was limited to those from this universe that mailed back their forms.

The five treatments were the Census 2000 Skip To Instruction, Go To Instruction, (Go To) Reverse Print Instruction, (Go To) Prevention Instruction and the (Go To) Detection Instruction, all of which were incorporated into questionnaires having black print on a light orange background with orange trim features.

3.1. Treatments

3.1.1. The Census 2000 Skip To Instruction

Shown in Figure 2a, this instruction was used in the classroom experiment, and is exactly the same as the instruction used on the Census 2000 long form.

3.1.2. The Go To Instruction

Shown in Figure 2b, this instruction is like the Census 2000 instruction in all respects, except that the words ''skip to'' have been changed to ''go to.''

3.1.3. The (Go To) Reverse Print Instruction

Shown in Figure 2c, this instruction is like the Go To instruction, except that the words ‘‘Go to’’ have been changed from normal print (black lettering on a light orange background in the census, but depicted as gray here) to reverse print (orange lettering in the census, but gray here, on a black background).

3.1.4. The (Go To) Prevention Instruction

Shown in Figure 2d, this is a modification of the prevention branching instruction from the classroom experiment, with ‘‘skip to’’ changed to ‘‘go to.’’ A bold arrow was placed between the check box and the branching instruction to make the connection between them. Also, the number of ‘‘attention’’ instructions was dramatically reduced. Finally, the language of the ‘‘training instruction’’ was simplified from what it had been in the classroom experiment.

3.1.5. The (Go To) Detection Instruction

Shown in Figure 2e, this is a modification of the detection instruction from the classroom experiment, with ‘‘skip to’’ 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.

3.2. Implementation procedures

The questionnaires in this experiment received *very nearly* the same implementation procedure as other questionnaires in Census 2000. The questionnaires were mailed out according to the Census 2000 schedule, with every sampled address mailed an advance letter, a questionnaire, and a follow-up postcard. One difference between the experimental and the census procedures, however, was that the experimental questionnaires were mailed back to the National Processing Office (in Jeffersonville, Indiana) rather than the nearest geographically designated processing office. Consequently, the color of the return envelope was changed from white to light orange to facilitate its being identified and therefore properly reaching the National Processing Office during the onslaught of census mail returns.

3.3. Analytic procedures

3.3.1. Calculating mail response rates

Households that returned duplicate forms were excluded from the calculation of the mail response rate (two cases), as were households who did not return a form, but who were subsequently labeled as undeliverable as addressed in the mail out file (2,834 cases). It was assumed in the latter case that the household was correctly labeled as nonexistent or vacant. However, households that were identified as undeliverable as addressed in the mail out 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 mail out file.

Nonresponse, then, was defined as any remaining household in the mail out 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. Response was defined as households from whom a nonblank form from the mail out universe was received. The aggregate totals for all responses and nonresponses were established and then the total number of responses was divided by the total number of responses and nonresponses to yield the mail response rate.

3.3.2. Calculating error rates

To control for differences in the number of questions that respondents answered, our analysis was limited to the questions for Person 1. There were 92 questions asked of Person 1. Of these, 19 contained branching instructions. Branching error rates were calculated for questions that had branching instructions (because only their designs differed between form types) and those questions that had valid responses (because only then was it evident whether a respondent should branch or not).

An error of commission occurred if a respondent chose a response with a branching instruction, but failed to branch correctly. The commission error rate is equal to the sum of those who failed to branch correctly (i.e., they answered the next question) divided by the sum of those who chose a response with a branching instruction (i.e., those who had the opportunity to make an error of commission).

An error of omission occurred if a respondent chose a response without a branching instruction, but instead of answering the next question, he or she branched. The omission error rate is equal to the sum of those who erroneously branched (i.e., they did not answer the next question) divided by the sum of those who chose a response without a branching instruction (i.e., those who had the opportunity to make an error of omission).

Commission and omission opportunities, errors, and rates were calculated by respondent, by question within a treatment, across all questions, and across all treatments.

3.3.3. 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.

4. Results

4.1. Response rates

Response rates for the five treatment groups varied significantly between the high and low coverage areas, averaging 66.7 percent for the former and only 48.6 percent for the latter

(Table 1). This difference of about 18 percentage points is not surprising inasmuch as the strata differ significantly with regard to the characteristics of residents. Respondents that mailed back their forms from the low coverage areas are about six times as likely as high coverage respondents to be nonwhite (53.8 percent vs 8.8 percent) and/or Hispanic (23.7 vs 4.25 percent). In addition, they are three times as likely to speak a non-English language at home (28.7 percent vs 9.6 percent) and more than twice as likely never to have graduated from high school (29.4 percent vs 13.1 percent). These characteristics have often been associated with lower response rates for mail surveys (Dillman 2000).

However, differences in response rates across treatment groups tend to be small or non-existent. Comparisons of the Skip To and Go To control groups, and all Go To treatment groups with one another did not reveal any significant differences.

4.2. *Commission error rates*

Significant differences existed for most of the treatment comparisons of commission error rates. It can be seen in Table 2 that the average commission error rate for the 19 branching items contained in the Census 2000 skip form (Treatment 1) was 19.7 percent. As expected, the rate was substantially higher in the low coverage areas (26.9 percent compared to 18.6 percent). Rates also varied substantially by item (not shown in tabular form), with the overall rates ranging from 1.9 to 79.2 percent.

The overall commission error rate for the Go To comparison (Treatment 2) was not

Table 1. Response rates by treatment

Treatment	Weighted national total (%)	Weighted *n* (in 000s)	High coverage areas (%)	Unweighted *n*	Low coverage areas (%)	Unweighted *n*
1. Census 2000 Skip To	63.9	(12,630)	67.5	(2,377)	48.5	(2,321)
2. Go To Control	64.3	(12,520)	67.8	(2,355)	49.5	(2,307)
3. (Go To) Reverse Print	61.8	(12,650)	64.9	(2,387)	48.1	(2,304)
4. (Go To) Prevention	63.1	(12,540)	66.7	(2,365)	47.6	(2,288)
5. (Go To) Detection	63.3	(12,660)	66.4	(2,388)	49.3	(2,305)
Statistical comparisons						
1 vs 2	n.s.		n.s.		n.s.	
2 vs 3	n.s.		n.s.		n.s.	
2 vs 4	n.s.		n.s.		n.s.	
2 vs 5	n.s.		n.s.		n.s.	
3 vs 4	n.s.		n.s.		n.s.	
3 vs 5	n.s.		n.s.		n.s.	
4 vs 5	n.s.		n.s.		n.s.	

statistically different from the Skip To treatment. Because all of the remaining treatment groups used the words ''Go to,'' Treatment 2 becomes the control group for the remaining comparisons. Table 2 reveals that all three of the remaining treatment groups achieved significantly lower commission error rates, 17.9 percent for Reverse Print, 14.7 percent for the Prevention Treatment, and 13.5 percent for the Detection Treatment. The Detection Treatment, which had the lowest rate, was significantly lower than each of the other treatment groups, with two exceptions, the comparison with the Prevention Treatment in the high coverage areas and at the national level. The Detection Treatment lowered the commission rate by about one-third for the strata as well as overall.

4.3. Omission error rates

Shown in Table 3, a somewhat different pattern emerges for the omission error rates. There were no significant differences between the standard Skip To form used in Census 2000 (Treatment 1), and the Go To (Treatment 2) version. In addition, the only treatment that significantly reduced the omission error rates was the Detection Treatment, which did so nationally and in high coverage areas. In contrast, the omission error rates for the Reverse Print and Prevention Treatments were significantly higher than the Go To Control Treatment for all groups. Thus, only the Detection Treatment significantly reduced both the commission and omission error rates.

Table 2. Commission error rates for all Census long-form items with branching instructions

Treatment	Weighted national total (%)	Weighted number of commission opportunities (in 000s)	High coverage areas (%)	Weighted number of commission opportunities (in 000s)	Low coverage areas (%)	Weighted number of commission opportunities (in 000s)
1. Census 2000 Skip To	19.7	(62,163)	18.6	(53,473)	26.9	(8,694)
2. Go To Control	20.8	(62,947)	20.0	(54,105)	25.4	(8,843)
3. (Go To) Reverse Print	17.9	(60,803)	16.7	(52,140)	24.9	(8,664)
4. (Go To) Prevention	14.7	(61,046)	13.6	(52,472)	21.7	(8,574)
5. (Go To) Detection	13.5	(62,146)	12.7	(53,699)	18.6	(8,448)
Statistical comparisons						
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 3. Omission error rates for all Census long-form items with branching instructions

Treatment	Weighted national total (%)	Weighted number of omission opportunities (in 000s)	High coverage areas (%)	Weighted number of omission opportunities (in 000s)	Low coverage areas (%)	Weighted number of omission opportunities (in 000s)
1. Census 2000 Skip To	5.0	(40,079)	4.8	(34,475)	6.5	(5,605)
2. Go To Control	5.4	(40,443)	5.2	(34,635)	6.3	(5,808)
3. (Go To) Reverse Print	7.6	(39,248)	7.3	(33,875)	9.1	(5,373)
4. (Go To) Prevention	7.0	(39,044)	6.7	(33,624)	9.4	(5,420)
5. (Go To) Detection	4.0	(39,451)	3.7	(34,030)	6.2	(5,421)
Statistical comparisons						
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	$p<.01$		$p<.01$		$p<.01$	

5. Discussion

5.1. Response rates

We hypothesized that changes in the branching instructions would have little effect on the response rates, and the results confirmed this (Table 1). 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.

5.2. Commission error rates

There is only one change between the Skip To and Go To treatments (the verbal instruction was changed from ‘‘Skip to’’ to ‘‘Go to’’); therefore, the effect of this change is controlled for by design. However, several distinct changes were introduced into each of the next three treatments (Reverse Print, Prevention, and Detection), making it impossible to disentangle with certainty the effects of any single manipulation. Despite this limitation, however, the treatments build upon one another in such a way as to be highly suggestive regarding the individual manipulations and languages, as discussed below.

Figure 3 shows that, except for the Go To version, the commission error rate (for all areas) declines across the treatments. It was originally hypothesized that changing the

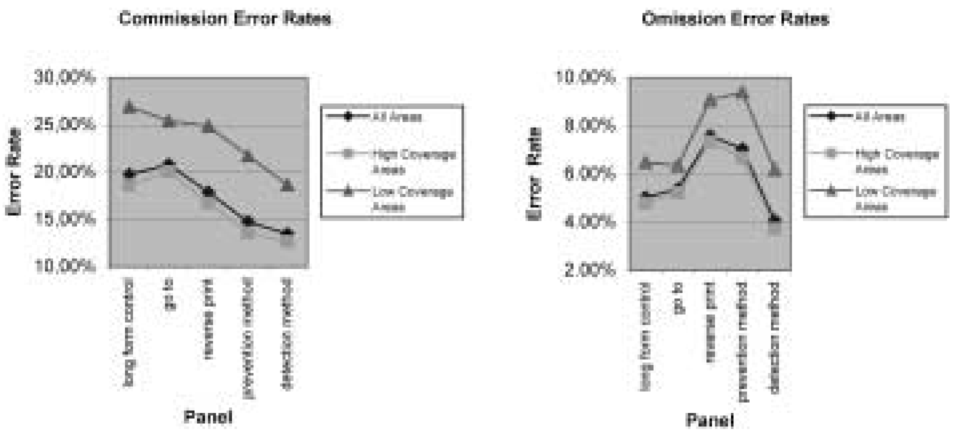


Fig. 3. Error rates by treatment

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. The similar rates of commission errors between these two treatments support this hypothesis. The lesson to be drawn from this finding, which in our experience has not necessarily been followed in practice, is that no amount of rewording is going to help if the problem is that respondents are *not* reading the information in the first place. Therefore, one should make certain that information is being read first before attempts are made to reword it.

The fact that the commission error rate decreases across the Go To, Reverse Print, Prevention and Detection treatments suggests that the changes made from one design to the next improved respondents' perception and comprehension of the instruction. We hypothesized that the Prevention and Detection treatments would have this effect. However, the theoretical arguments were ambiguous concerning what to expect from using reverse print because of the expected opposite effects of greater contrast versus the figure-ground change. Although the Reverse-Print treatment led to a reduction in the commission error rates, this reduction was mediocre in comparison to the Prevention and Detection treatments. This finding implies that there are probably better ways to attract respondents' attention to information on a questionnaire than using reverse print. It may be that respondents get used to reading information in a particular figure-ground (black against light orange in the case of the actual census questionnaires). As a result, they come to expect that the information they should pay attention to will be black against light orange too. When the instruction is reverse printed, it may look so different that respondents have a greater tendency to disregard it. According to psychologists a number of perceptual principles guide our understanding and interpretation of visual information. The Grouping Law of Similarity states that we tend to see similar information as belonging together. Thus reverse printing information may be an example of not using the Grouping Law of Similarity, and in particular the visual element of color in a beneficial way (Wallschlaeger and Busic-Snyder 1992).

The improvement in performance between both the Prevention and Detection treatments over the other treatments is likely due to making respondents more aware of the branching instructions, i.e., making the branching instructions more visible. However, the added improvement, which the Detection treatment displayed over the Prevention treatment in the low coverage area may be due to the feedback mechanism. It appears that the redundant phrasing in this mechanism may have effectively helped these particular respondents to self-correct their mistakes.

5.3. Omission error rates

Figure 3 shows that the errors of omission decreased for the Detection treatment in the census experiment, but increased for every other method. It would seem that unlike the Reverse Print and Prevention treatments, which were focused on making the branching instruction more visible, the Detection treatment successfully grouped information in the census experiment so that respondents were less likely to erroneously associate a branching instruction with a nonbranching response option. The detection method is the only method tested that labeled all response choices with symbols as well as words. This dual labeling might have helped respondents who did not understand that they were expected to continue to the next question unless otherwise instructed. It would seem that the left-hand arrow served the purpose intended here – to lead respondents' attention away from the branching instruction when they chose a response option without such an instruction.

The implication of this finding is that grouping information correctly may be tantamount to respondents reading and understanding it correctly. If so, this is an example of the power of the Grouping Law of Proximity (Wallschlaeger and Busic-Snyder 1992), that is, of manipulating the visual element of location.

These results suggest that at the same time as 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 or 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).'' For example, the errors of omission soared to 17.6 percent in the case of the citizen question in the Detection treatment of the census experiment, whereas it averaged 4.1 percent across the rest of the questions using this same method. The feedback mechanism for the citizen question was ''(If born outside or not a citizen of the United States)'' whereas it tended to be ''(If Yes)'' or ''(If No)'' for the rest of the questions.

5.4. Questionnaire design guidance

An important finding to come out of this research is that we must be careful what information we use to guide us when designing questionnaires. Early on, when questionnaire design experts perused the questionnaires they had negative reactions to the use of the arrows in the Detection treatment, claiming that the arrows looked confusing. Respondents in cognitive interviews often said the same (Dillman et al. 1999; Redline and Crowley

1999), and even a debriefing of respondents who participated in the classroom experiment revealed that some felt the arrows were confusing (Redline et al. 1999). These findings suggest that expert reviewers and respondents perceive themselves as being confused, when the outcome suggests otherwise. This may be occurring for one of two reasons. Pre-attentive processing involves the automatic registration of features at a global or holistic level, whereas attentive processing requires a detailed analysis of a field (Jenkins and Dillman 1997). Quickly perusing the questionnaire (that is, using pre-attentive processing), as the abovementioned questionnaire design experts had done may lead to a perception of confusion, whilst actually filling it out (using attentive processing proves otherwise) as demonstrated by the results of this experiment. And the other possibility is that respondents perceive themselves as being confused because the feedback mechanism is working – it is slowing them down and helping them to correct their mistakes. In reality, though, their performance is improved.

5.5. Question and respondent effects

Redline and Dillman (2002) have proposed that respondents make mistakes navigating through a questionnaire both as a result of the characteristics of the questions they are answering (for example, questions that fall at the bottom of the page may lead to larger errors than those located elsewhere) and as a result of respondents' characteristics (for example, respondents with less education may make more mistakes). Wide variations existed in the error rates for individual questions. For example, on the Census 2000 Skip To form, commission errors ranged from a low of 1.9 percent for the active duty question (Question 20a) to a high of 79.2 percent for the age filter question (Question 18). Differences in the error rates across the different versions of the branching instructions, across the different question types, and between respondent types (that is, between the coverage areas) provide evidence in support of both propositions. The conclusion to be drawn is that this is a highly complex system under investigation, the effects of which it is clearly going to take time to explore and explain well.

5.6. 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 as to 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, a surprising finding to come out of this research is that the error rates are either the same or higher in the census as compared with 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 screener questions contain branching instructions, which clearly tell respondents to branch over the follow-on questions;
- the follow-on questions contain contrary cues, which ought to keep respondents from answering them;
- survey practitioners perceive the questions as logically connected, and therefore think that respondents will too.

Consequently, not only may it be a good idea to improve upon the branching instruction, but also it may pay to recognize that respondents do not necessarily understand the basic logic of the questionnaire and the questions themselves, and that any steps taken to improve their understanding of this may help to reduce navigational errors as well.

6. Conclusion and Suggestions for Future Research

This article provides 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, symbolic, and verbal languages that comprise branching instructions influences significantly whether those instructions are followed. Thus, our general hypothesis that visual language affects how respondents navigate and complete questionnaires is supported.

As hypothesized, respondents were not affected by the instruction variation ''go to'' versus ''skip to.'' Together with the other findings, this suggests that rewording is not the solution if the problem is that respondents are not reading the information in the first place. Therefore, we need to make certain information is being read before we go to the trouble of rewording it. In addition, respondents were more likely to misread the reverse-printed instructions, suggesting that reverse-printed material may be more effective when a reader is actively searching for information as opposed to passively reading it. The implication of these findings is that we need to continue to gain expertise regarding the nonverbal (numeric, symbolic, and graphic) languages of a questionnaire, for the nonverbal combines with the verbal to affect reading comprehension – that is, what respondents read (or do not read), the order in which they read it, and their consequent interpretation of what they read.

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

However, an important finding to come out of this research is that relying on the verbal assessments of either respondents or expert questionnaire designers rather than the performance measures from the experiments would have been misleading and detrimental because although error rates were the least in the Detection treatment, questionnaire design experts, respondents in cognitive interviews, and even debriefings of respondents who participated in the classroom experiment revealed that many thought the arrows of the detection method were confusing. The conclusion to be drawn from this is that what people say and what people do are not necessarily one and the same, and that therefore we must exercise care regarding what information we use to guide us when designing questionnaires. Further research is necessary to determine when verbal reports are reliable and when they are not.

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 for combining the Prevention and Detection treatments in order to harness the theoretical power that certain aspects of each seem to offer for improving performance with branching instructions, along with disentangling the contributing effects of individual manipulations of the Detection treatment, since it is now clear that the manipulations are successful in combination.

The classroom experiment used college students who have a higher level of education than many of the respondents in the census experiment. Also, the two experiments differed in where the form filling task took place (classroom versus field), the substance of the questions (life styles versus demographic questions) and the manner in which the questions were connected to one another (independent versus dependent). The fact that the overall pattern of commission error rates is similar in the two experiments despite these differences is robust evidence that the instructions differ in their ability to affect respondent performance.

Having said that, however, the absolute error rates within a treatment are either the same or higher in the census as compared with the classroom. This suggests that the general population has a greater tendency to answer questions that do not apply to them under field conditions, despite the questions containing cues to the contrary and branching instructions. This is further evidence that the general population does not understand the basic intent of the questions or the fact that not all questions apply to them. Thus, not only may it be a good idea to improve upon the branching instructions, but it may pay to recognize that respondents do not necessarily understand the basic logic of the questionnaire and the questions themselves, and that any steps taken to improve their understanding of the questions may help to reduce navigational errors as well.

Finally, we have shown that respondents extract meaning from more than the verbal language of the questionnaire. Now that the major thesis of our research has been borne out with branching instructions, it needs to be systematically extended to other areas of the questionnaire, like the questions themselves, and it is important to bear in mind that these issues extend to Web questionnaires as well, which also require respondents to extract information through a visual mode.

7. References

Dillman, D. (2000). Mail and Internet Surveys: The Tailored Design Method. New York: John Wiley and Sons.

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.

Featherston, F. and Moy, L. (1990). Item Nonresponse in Mail Surveys. Paper Presented at the International Conference on Measurement Errors in Surveys, Tucson, Arizona.

Foster, J. (1979). The Use of Visual Cues in Text. Processing of Visible Language, 1, 189–201.

Hartley, J. (1981). Eighty Ways of Improving Instructional Text. IEEE Transactions of Professional Communication, 24, 17–27.

Jenkins, C. and Dillman, D. (1997). Towards a Theory of Self-Administered Questionnaire Design. In Survey Measurement and Process Quality (L. Lyberg, P. Biemer, M. Collins, C. Dippo, E. De Leeuw, N. Schwarz, and D. Trewin (eds)). New York: Wiley-Interscience.

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

Messmer, D.J. and Seymour, D.T. (1982). The Effects of Branching on Item Nonresponse. Public Opinion Quarterly, 46, 270–277.

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

Redline, C. and Dillman, D. (2002). The Influence of Alternative Visual Designs of Respondents' Performance with Branching Instructions in Self-Administered Questionnaires. In Survey Nonresponse, (Groves, R., Dillman, D., Eltinge, E., and Little, R. (eds.)). New York: John Wiley and Sons, Inc.

Redline, C. and Lankford, C. (2001). Eye-Movement Analysis: A New Tool for Evaluating the Design of Visually Administered Instruments (paper and Web), Proceedings of the American Statistical Association, Section on Survey Research Methods.

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

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

Turner, C.F., Lessler, J.T., George, B.J., Hubbard, M.L., and Witt, M.B. (1992). Effects of Mode of Administration and Wording on Data Quality. In Survey Measurement of Drug Use Methodological Studies, C.F. Turner, J.T. Lessler, and J.C. Gfroerer (eds). Washington, DC: National Institute of Drug Abuse, U.S. Department of Health and Human Services, 221–243.

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 Research Evaluation Division, December 22.

Received June 2002
Revised September 2003