How Visual Grouping Influences Answers to Internet Surveys*

Jolene D. Smyth, Don A. Dillman,
Leah Melani Christian, and Michael J. Stern
Washington State University

INTRODUCTION

Each question on a self-administered questionnaire can be viewed as a group of visually presented content that consists of the query, any instructions, response choices (unless open-ended), and spaces where answers are to be marked. Survey researchers have proposed that questionnaires be designed in a manner that clearly shows each question as a grouping distinct from all other questions and that each question group be presented in a way that the respondent can infer the order in which questions and their sub-parts are to be processed and answered (Dillman, 2000). Thus, the questionnaire can be viewed as a sequence of information that is divided into interconnected groupings and sub-groupings of question content. However, relatively little research has been done to conceptualize and test how grouping information is communicated and whether changes in grouping affect measurement or other response behaviors.

Web surveys are of particular interest because they provide far more potential than paper questionnaires for manipulating information in ways that visually change how it is interpreted by respondents. On the Internet it is easy and inexpensive to manipulate the visual aspects of pages in various ways. In addition, new question formats (e.g. drop down menus) and prescribed uses of symbols (radio buttons vs. html boxes), produce visual variations that may contribute to the occurrence of unintended grouping effects, which may in turn affect answers.

Our purpose in this paper is to report results from a series of experimental manipulations for three types of questions to examine whether alternative visual grouping influences respondent answers to survey questions. They include: 1) the effects of providing headings for subgrouping of response options, 2) alternative procedures for presenting long lists of answer choices to respondents, and 3) the differential location of symbols to define requests for dates. The three question formats subjected to experimental testing were identified as presenting significant difficulties to pretest respondents during the evaluation of a web prototype of the NSF sponsored National Survey of Earned Doctorates (Altheimer and Dillman, 2002). This research was undertaken to identify possible solutions to those problems as well as contribute to our understanding of how visual design decisions influence respondent answers through grouping processes.

* A more detailed version of this paper, complete with figures that provide exact question wording and visual layout, is available from the Social and Economic Sciences Research Center (Technical Report #04-023) at Washington State University or at the following web address: http://survey.sesrc.wsu.edu/dillman/papers.htm.

BACKGROUND

Adherence to the Rules of Communication

Respondents to surveys follow rules or maxims of communication (Schwarz 1996). In other words, they approach the survey instrument as if they are in a conversation, with the instrument representing the researcher's contribution to the conversation. Within this context Schwarz highlights how apparently formal features (e.g. graphical layout) of the questionnaire from the researcher's perspective are important in the answering process because they communicate to the respondent what is expected of them. The importance of these formal features is magnified when the respondent is unsure what is being asked of them or how they are expected to answer. In these situations respondents are more likely to take their cues from design features of the questionnaire than when the researcher's expectations are more clear (Schwarz 1996).

Within the framework of respondents as cooperative communicators grouping is highlighted as an important formal feature of questionnaire construction that communicates expectations to the respondent. It can help respondents understand the intent of the question, response options, and answer spaces and thus reduce their likelihood of committing errors. For example Schwarz and Hippler (1992) found that respondents gave different answers to two questions (asking about marital and general satisfaction) when the questions were grouped together by placing them within a single box, as opposed to when they were presented separately in two boxes. Specifically, levels of general satisfaction and marital satisfaction were less correlated when both questions were presented together in a box (as a group) than when they were presented separately (in two boxes). This finding indicates that the grouping of the questions affected how respondents interpreted them and subsequently what responses they gave.

Visual Processing as a Basis for Grouping

When respondents first look at a questionnaire they use preattentive processing (Neisser 1967) to quickly take in the whole scene and make sense of the information presented (Jenkins and Dillman 1997). At this broad level of processing all objects in the field of vision are competing for the respondent's attention (Neisser 1967). It is during this stage that certain features of the questionnaire (e.g. the number one or a bold sign saying "start here") are likely to capture respondent attention. Attentive processing involves respondents choosing a part of the questionnaire to focus on and then shifting their attention to another part, moving through the available information until the survey is completed. According to Neisser (1967) these attentive acts are "carried out in the context of the more global properties already established at the preattentive level" (p.

90). Thus, survey designers can help direct the respondent at both the preattentive and attentive processing levels. For example, section headings and boxes encompassing questions are sometimes provided to help respondents group questions at the preattentive processing level (e.g. Dillman 2000: 397), whereas the content of individual items—the query, any instructions, and answer choices—are usually grouped in a consistent format within the question for consideration at the attentive stage of processing. Respondents use these groupings and sub-groupings as tools to navigate through the questionnaire (Dillman, 2000).

The Use of Visual Language to Achieve Grouping

A number of different methods are available for survey designers to use in their efforts to influence how respondents comprehend questionnaires and the individual items they contain. The most obvious is to manipulate the verbal language, or words, used to communicate with the respondent. However, research has shown that manipulating verbal language is only one way to convey information as respondents also rely on nonverbal languages to determine meaning (Redline and Dillman, 2002). Nonverbal languages include graphical language (font size, color, brightness, spacing, etc), numerical language (the use of numbers to suggest order), and/or symbolic language (i.e. the use of culturally prescribed symbols such as arrows to direct movement through questions). Nonverbal languages are used in conjunction with verbal language to communicate certain meanings to the respondent such as where to start, where to proceed next and how to process a specific question. In addition to its direct affects, graphical language also serves as the conduit through which the other languages are transmitted. In other words, graphical manipulations (changes in the size, shape, location, spatial arrangement, color, brightness, contrast, and figure/ground composition) can influence the way verbal, numeric, or symbolic languages are perceived and interpreted (Jenkins and Dillman 1997). Thus, they play a crucial role in guiding respondents through the answering process. However, without careful attention to detail nonverbal language can work in opposition to verbal language cues and lead the respondent to complete the survey in ways unintended by the researcher, thus introducing response errors. One example is when respondents do not understand and thus fail to follow branching instructions, answering the questions in incorrect orders or failing to respond to applicable questions.

Nonverbal language can be used effectively to create desired groupings and sub-groupings to simplify the answering process. The principles of proximity, similarity, and pragnanz are three relevant pattern recognition concepts from Gestalt psychology which have been described by Palmer (1999) and applied to survey design by Jenkins and Dillman (1997). These Gestalt principles suggest several ways to group information in questionnaires. One way is through the use of space on the instrument. For example, using greater space between questions than between the stem of a query and the accompanying response options creates grouping based on proximity and clarifies the boundaries between questions. Another way of establishing grouping is through similarity (or contrast). Through graphical manipulations, one may make all question stems bold or in a larger font, while keeping response options smaller, thus establishing subgrouping. Similarity can also be achieved through orientation. For example, items that are oriented horizontally on the page might appear to belong to one group while items that are oriented vertically appear to belong to another (Palmer 1999: 258).

Web Surveys

As web surveys grow in popularity and frequency it is important to understand how grouping and subgrouping can be conveyed in them as well as some of the additional abilities and challenges that web survey designers confront. To the extent that they rely on visual presentation of material, web surveys may be very similar to paper surveys—correct completion of them depends on the graphical presentation of verbal as well as nonverbal languages. Experiments have shown that varying the layout of multiple answer questions (Dillman, Smyth, Christian and Stern, 2003) and ordinal scale questions (Christian, 2003) produces similar differences in paper and web surveys. However, web surveys also differ from paper surveys in a number of important ways. For example, one fundamental difference is the ability to affect grouping via the number of questions included on any one screen (Best and Krueger, 2004). Whereas it is not plausible to construct a paper survey with one question per page to designate grouping, it is possible to construct a web survey in this manner, referred to as page-by-page construction Web surveys also differ from paper surveys with respect to page orientation (i.e. the vertical orientation of an 8.5 x 11 inch paper survey compared to the horizontal orientation of the typical computer screen) as well as options for presenting response options (i.e. the use of drop-down menus and the decision between using html boxes and radio buttons).

The avoidance of errors is important for the sake of measurement in all surveys, but it is especially important on web surveys where oftentimes respondents will be informed of their error and asked to correct it through an error message. Although error messages might help the respondent provide an acceptable answer, they may also promote a certain level of frustration that could result in respondents failing to complete the survey (Altheimer and Dillman 2002). This is especially true if respondents receive several such messages and still cannot figure out what is expected of them. Finally, clear grouping can indirectly affect respondents' frustration levels by reducing the amount of time it takes to answer questions and thus shortening the perceived length and personal costs of completing the survey.

Problem Question Formats in the Earned Doctorate Survey Web Prototype

In 2001 a longstanding paper questionnaire survey conducted by the National Science Foundation, The Earned Doctorate Survey, was converted to a web survey format. Attempts are made each year to get every person who finishes a doctoral degree at a U.S. university to complete this survey. Because it was anticipated that this survey would be conducted for a number of years as a mixed-mode survey, and because of the importance of maintaining trend lines, it was deemed important that attention be given to achieving mode comparability. Therefore, cognitive interviews were conducted to evaluate a pilot version of the web survey. These revealed significant problems on three question formats. For all three questions, problems appeared to stem from grouping and subgrouping issues (Altheimer and Dillman 2002). Alternative versions of these questions are experimentally compared in this study to better understand how grouping processes may influence respondent answers.

The first question included a set of response categories that appeared to be visually divided into two groups through the insertion of descriptive headings, but respondents were expected to choose only one answer. In the cognitive interviews respondents exhibited a tendency to try to choose at least one answer from each group despite the fact that the radio button format allowed only one answer to be selected overall. The second question was a request for respondents to choose a "field of study" from a list of fields divided into categories and subcategories. On the web survey this display, taken directly from the paper questionnaire, required the use of two screens. Respondents had difficulty toggling between the two screens to find their appropriate field of study. The third question was simply a request to report month and year that the respondent started graduate school. To emphasize that two digits were to be used to report month and four to report year, the symbol "MM YYYY" was placed after the year box. However, respondents made various errors including among others the use of letters for the month or trying to place two digits in each box. Respondent difficulties to each of these pilot study questions led to error screens that produced additional respondent frustration (Altheimer and Dillman 2002).

The cognitive interviews revealed the need to better understand how the grouping of information is communicated to respondents and whether some versions are more effective than others in getting respondents to answer questions accurately. In the present study we adapted each of these questions to topics and formats that could be evaluated in web surveys conducted of undergraduate students. We report here results from a series of two web surveys, each with up to four comparisons.

PROCEDURES

Two web survey experiments were conducted, with results from the first experiment being used to guide the design of the second one. The first web survey, which consisted of 21 questions, was conducted in the spring of 2003. The survey was designed to assess the undergraduate experience at Washington State University (WSU). It used a page-by-page design and questions were presented in black text against a colored background with answer spaces appearing in white so as to provide contrast between the answer spaces and the background. All screens were constructed with HTML tables using proportional widths in order to maintain the visual aspect of the screen regardless of individual user window sizes. In addition, font size and style were automatically adjusted using Cascading Style Sheets to accommodate differing user screen resolutions. The sample consisted of 3,004 randomly selected WSU undergraduate students who were registered for courses on the Pullman campus during the spring 2003 semester. Of the 3,004 students in the sample 1,591 completed the survey for a response rate of 53 percent.

The second web survey, consisting of 25 questions, was conducted in the fall of 2003. This survey was designed to assess students' experiences both on and off campus in the Pullman, Washington area. Similarly to the first survey, this one was designed using page-by-page construction with questions presented in black text against a colored background. Answer spaces appeared in white. Similar precautions were taken to ensure that the appearance of the questions would not be manipulated in important ways by different computer configurations. The sample consisted of 3,045 randomly selected WSU undergraduate students who were registered for courses on the Pullman campus during the fall of 2003. The response rate for this survey was 56 percent or 1,705 students.

Sampled students were contacted by postal mail and asked to go to the web and complete the questionnaire. With the initial contact letter respondents received a two-dollar incentive. E-mail follow-ups to provide a hotlink and two additional postal contacts were made. To gain entry to the survey instrument students were required to enter their own personal access code, which was provided to them in the first contact letter and all subsequent contacts. Access codes were used to ensure that only individuals in the sample could participate and that they could only complete the survey once. A random number generator was used to assign one of the versions to each respondent when they entered the survey. For both surveys client side paradata were collected (Heerwegh and Loosveldt 2002; Stern et al. 2004). Of specific value to the questions being explored in this paper is the time respondents spent on each question.

THE INFLUENCE OF CATEGORY HEADINGS ON GROUPING

Respondents to the NSF Earned Doctorate Survey were asked to indicate their immediate postgraduate plans by choosing one answer from among seven choices. In the NSF mail questionnaire these categories were visually separated with four of them being placed under a general heading of "further training or study" and the remaining three categories being placed under the heading of "career employment." This format was retained for the pilot web survey where it was observed that some respondents tried

to provide answers under each heading. However, the use of radio buttons for this question meant that some respondents unintentionally erased their answer to the choice they made under the first general category when they chose another under the second general category without being aware of having made the change. The grouping issue addressed by this question is whether spacing differences in combination with the use of words (expressed as headings) influence people's answer choices. Our general hypothesis is that when response choices are presented to respondents as separate groups they are more likely to choose answers from both groups.

The NSF question requesting immediate postgraduate plans was operationalized in our first survey by asking students to indicate which of six options described the benefits of the student recreation center. Four slightly varying versions of this item were constructed. The fourth is not reported here because it had no significant effects. The remaining three treatments were as follows (See the footnote on page 1 for information on how to acquire screen shots of all the experiments reported here):

Version 1: An underlined heading was placed above each of two subsets of three response options. One was labeled "Health Benefits" and the other was labeled "Academic Benefits." In addition, there was one line of space left blank between the two subsets. This version emulates the original NSF question format.

Version 2: The same question and groupings were presented, but a word instruction stating "Please select the best answer" was added to attempt to override any grouping effect.

Version 3: All six choices were placed in a single vertical line with no indication of sub-grouping (no headings and no additional spacing between groups).

The experiment included in the second survey provided a different question topic and somewhat different approaches to grouping, as well as efforts to override any effects. The topic of this operationalization was sources of financial support. The subgrouping of the eight response options included the presentation of half of the options in a second column (i.e. a double-banking format). In other words, the sub-groups were aligned with each other horizontally instead of vertically as in the first survey. The two relevant treatments were as follows:

Version 1: Headings reading "Financial Aid" and "Other Sources" were placed above two double banked sub-groups of four response options each.

Version 2: The double banking was retained, but the headings were removed.

We expect the headings to have less effect in the second survey; the groupings are less visually prominent and are overshadowed by the use of the two column format. In addition, it has been suggested that double-banking leads to a tendency for the second column not to be seen as it is often placed outside of the foveal view of a respondent.

Results for Category Heading Effects

Results from the first set of experiments make it clear that the use of the headings influenced answer choices, as expected (see Table 1). Respondents to version 1 (headings and separation) of the first survey not only chose more response categories than the respondents to version 3 (no headings or separation), but were more likely to choose at least one answer from each of the sub-groupings (70.2% vs. 40.9%). These results clearly suggest that the use of headings and accompanying separation to establish two visual sub-groupings influence responses.

The addition of the instruction to "Please select the best answer" in an attempt to override the visual grouping effect (version 2) had a mixed effect. Respondents to the version with the instruction (version 2) were only slightly less likely to choose an answer in each sub-grouping than were those who completed the version without the instruction (version 1) (chi square = 1.630, p = .202). However, the mean number of options checked was significantly less for those who had the additional instruction (version 2) than for any of the other treatments. These results intimate that respondents may have been drawing information from two sources, the instruction to select the best answer (the lower mean) and the sub-groupings (the greater tendency to select from both groupings). These two sources of information in concert seemed to have indicated to the respondents that they were expected to select the best answer from each sub-group.

Table 1: The Effects of Category Headings on Vertical Grouping

Q18: What best describes the benefit of the Student Recreation Center (Survey #1)			
	V 1 Headings	**V 2 Headings/ Instruction**	**V 3 No Headings**
% Marked in Both Groups	70.2	66.0	40.9
Mean Number Checked	3.0	2.4	2.7
n	**435**	**438**	**367**

SIGNIFICANCE TESTS		
	Headings vs. No Headings	**Headings vs. Headings/ Instruction**
% Marked in Both Groups	X^2 = 62.36, p = .000*	X^2 = 1.630, p = .202
Mean Number Checked	t = 2.38, p = .018*	t = 5.41, p = .000*

The results from the second survey (Table 2) are somewhat different than those from the first. There was no significant difference between the percent of respondents who marked responses in the left, right, or both left and right groups across the versions with and without the

headings. In addition, there was no significant difference in the mean number of response options marked in the left, right, or overall across the version with the heading and the version without. These findings are consistent with our expectation that the groupings are overshadowed by the use of the double-banking format.

Overall, these findings suggest that the headings and spatial separation play a larger role when the groups are aligned vertically as opposed to when they are aligned horizontally. They support the hypothesis that the grouping is overshadowed by the double-banking of the response options.

Table 2: The Effects of Category Headings on Horizontal Grouping

Q21: Have you received financial support from each of the following while attending WSU? (Survey #2)		
	V 1 Headings	**V 2 No Headings**
% Marked in Both Groups	64.6%	66.8%
Mean Number Checked	2.9	3.1
n	**393**	**446**
SIGNIFICANCE TESTS		
	Headings vs. No Headings	
% Marked in Both Groups	$X^2 = 0.38$, p = .538	
Mean Number Checked	t = -1.86, p = .064	

SELECTING A FIELD OF STUDY

The second problematic question in the NSF Earned Doctorate Internet Survey requested that respondents choose their Ph.D. Field of Study from a long list of possibilities, which because of the number of options (282 fields organized under 13 headings and 16 sub-headings) had to be displayed on two different screens. On the paper questionnaire these options were presented on a single page at the end of the survey and did not appear to cause significant problems for respondents. However, several difficulties were encountered by respondents in the web prototype. For example, when their desired fields were not available some respondents repeatedly toggled between screens trying to find an appropriate response option. The toggling was inefficient and frustrating to a number of these respondents, appearing to be a significant problem for them.

Several grouping issues emerged from these observations. One concern was to find a more efficient method of locating the desired response category. Drop-down menus emerged as a possibility, but seemed to present a potential problem if they were to be constructed in the three level manner used for the paper questionnaire (general area, sub-area, and specific field). If a respondent entered a general and then sub area field only to find that their specific field was not included he/she would have to initiate another hierarchical search through another path. Observations in cognitive interviews indicated that this was a realistic scenario (Altheimer and Dillman 2002). In fact, respondents searching for fields such as environmental science, criminal justice, and statistics searched under various general areas until they found the best or perhaps least objectionable answer.

Another possible solution to the grouping issues was to present the groups of fields as one completely alphabetized list (one step) rather than as alphabetized fields contained under subheadings (multiple steps). A final possibility was to pose the field of study question as an open-ended question, thus avoiding the grouping process altogether. These alternative possibilities posed two main difficulties: (1) making provision for the possibility that students had more than one field of study and (2) determining whether or not the structure of the question would encourage or discourage the reporting of multiple majors.

A two-step experimental design was used to address these issues. In the first experiment, respondents were asked to report their major. The following are the treatments:

Version 1: The question was followed by an open-ended answer space large enough for reporting more than one major. In this treatment grouped information was not presented to the respondent. Instead, determining and reporting an appropriate answer was left entirely to the respondent.

Version 2: A two-step drop down menu was used. The first menu listed general areas as used on the University web page. When a general area was clicked specific majors would appear in the second drop down menu. It was impractical to test the three stage drop down menus approach with our population of undergraduate students.

Version 3: An alphabetized list of all university majors was presented. Respondents used the scroll bar to proceed through the list. Radio buttons were used so that only one major could be selected.

Version 4: An alphabetized list organized by general areas (categorized in the same way as in version 2) was presented. Radio buttons were used so that only one major could be selected.

The second experiment differed from the first in that it was constructed to allow for the possibility of multiple majors. The treatments were the same as in the first experiment save for accommodations made to allow the second or subsequent majors to be reported. For example, the wording of the question was changed to ask the respondents to report their "major or majors." In addition, two sets of drop down menus were provided for the second version of this experiment. Answer spaces for versions three and four consisted of html boxes rather than radio buttons to allow for multiple selections as well. An inadvertent programming error resulted in one of the questions (version 3) retaining the original question wording (asking for "major" instead of "major or majors").

The test of these formats across the two experiments embraces several grouping issues. One of them is the efficiency of the search process, and whether respondents are able to locate and mark their answer(s) more quickly in some formats than others. Another involves grouping based upon one step versus a multiple-step process. In addition there is the issue of providing a complete group from which answers are to be selected versus depending upon respondent recall to provide an answer in an open-ended fashion. Thus, this experiment involved different methods of grouping verbal information (completely alphabetical vs. alphabetical within general areas) as well as graphical layout.

Results for Selecting a Field of Study

The most important issue with this question is which format for grouping the response options provides for the most efficient processing of the question and response options. Through the use of paradata the mean times spent on each version were calculated (See Table 3). Outliers were removed at two minutes. For the first survey the open ended version was completed the quickest with a mean response time of 9.6 seconds. This was followed by the alphabetical list (14.2 seconds) and then the alphabetical list under general headings (21.1 seconds). Due to a programming error no time data was available for the drop down version on this first survey. That data was available on the second survey, however.

On the second survey the mean time it took to complete the open ended version was lowest at 15.3 seconds.[1] This grouping scheme was followed by the drop down menus version which took on average about 34.4 seconds to complete. Responding to the alphabetical list version took a mean time of 50.0 seconds. When the response options were grouped alphabetically under general areas the mean response time was 57.3 seconds. The format that took the longest to respond to was the grouped/alphabetic format which most closely resembles the NSF format. In terms of time, the open ended version was the most efficient while the grouped/alphabetic version was the least efficient. However, the open ended version also required coding by the researchers and introduced the ability for the respondent to use unclear abbreviations, provide unsolicited explanation, and list as many majors as space allowed.

Another important finding regards the reporting of double majors. Whereas the first survey did not allow respondents to report double majors, except in the open ended version, the second was designed specifically to do so. Results indicate that on average across the four versions about 15 percent of respondents reported double majors, suggesting that the inability to do so in the first survey may have inconvenienced a significant number of respondents. As a matter of fact, paradata analysis indicates that only one percent of respondents to the open-ended version of the first survey changed their answer. In contrast, 10.5 percent of respondents to the alphabetical listing and 13.9 percent of respondents to the grouped/alphabetical version changed their answers. Of those changes on the alphabetical and grouped/alphabetical versions 63.6 percent appeared to be due to the lack of ability to report double majors. These respondents marked an answer and then attempted to mark another answer, but apparently upon realizing that their second answer voided their first they then went back and re-marked their original selection.

Table 3: Effects of Grouping on Time Spent Answering the Field of Study Question in the First Survey

	TIME (SECONDS)	
	SURVEY #1	SURVEY #2
Open-Ended	9.6	15.3
Drop-Down	----	34.4
Alphabetical	14.2	50.0
Grouped/Alpha	21.1	57.3
	TESTS OF SIGNIFICANCE	
Open Ended vs. Drop-Down	----	t = 15.17, p = .000
Open-Ended vs. Alphabetical	t = -5.43, p = .000	t = -27.12, p = .000
Open-Ended vs. Grouped/Alpha	t = -12.31, p = .000	t = -32.41, p = .000
Drop-Down vs. Alphabetical	----	t = -10.13, p = .000
Drop-Down vs. Grouped/Alpha	----	t = -14.85, p = .000
Alphabetical vs. Grouped/Alpha	t = =6.06, p = .000	t = -4.63, p = .000

REQUEST FOR MONTH AND YEAR OF AN ACTIVITY

The third question format evaluated in this study was a simple request for respondents to record the month and year they completed their degree, using two digits for the month and four digits for the year. In the NSF web survey prototype several respondents completed this question inaccurately—either entering letters for a month or only two digits for the year—and received an error message. As a result of these difficulties the NSF question that asked for the month and year of degree completion was converted to a question asking undergraduates for the month and year that they began their studies at WSU. The

[1] The mean times for the second survey are consistently greater than for the first because they are page-elapse times whereas the mean times for the first refer to how long it took to check an answer independently of the time it took the page to load. The page-elapse times are used for the second survey so that the drop down version could be included. The important thing here is not the exact time it took for each version, but the consistent pattern of times relative to one another across the two surveys.

following treatments were designed to ascertain how best to get respondents to use the correct reporting format:

Version 1: This version emulates the original NSF format with the symbols "MM YYYY" located to the right of both answer boxes.

Version 2: The same answer boxes are used in this version, but the symbols are moved directly below the boxes such that the "MM" is located below the month box and the "YYYY" is located below the year box.

These specific treatments were designed because it was hypothesized that the abovementioned errors might result from the grouping of the month and year boxes together, followed by the grouping of the symbols "MM YYYY" together after the boxes. Thus, the symbol for the month was placed after the answer box for the year, separating the instruction from the answer space to which it was intended to apply. Similarly, the symbol for the year was placed after the symbol for the month, creating interruption between it and the answer space to which it was intended to apply. In addition, the "YYYY" may fall outside the respondents' foveal view. Version 2 was designed using the principle of proximity. We expect the closer grouping of the specific symbols with their respective answer boxes to result in a more clear understanding of expectations among respondents and as a result more respondents using the desired format.

Results for Month and Year of an Activity

Respondents who answered the date questions using a two-digit month and a four-digit year were regarded as using the "desired format." Results indicate that locating the symbols below the answer spaces as opposed to locating them to the right of the answer spaces resulted in a significantly larger proportion of respondents using the four-digit format for reporting the year on this question (94% vs. 89.9%; $X^2 = 5.93$, $p = .015$). As such, the version in which the symbols were placed below as opposed to at the right of the answer box produced more use of the desired format (90.6% vs. 88.5%), although the difference was not significant.

DISCUSSION AND CONCLUSIONS

In this paper we have investigated, experimentally, how certain aspects of the visual design and layout of questions influence how respondents process survey questions and provide answers in web surveys. Our focus is on understanding multiple aspects of the answering process, from whether visual layout changes people's answers to how best to communicate exactly what the designer wants so response errors, and corrections, are minimized. We are also interested in the efficiency of formulating answers. Thus, our interest is not only in obtaining accurate answers, but also in minimizing respondent burden and frustration.

The central issue explored in this paper is how different visual groupings of information affect how respondents arrive at answers—which designs promote accuracy, which are most efficient, and which will minimize the need for corrections.

Each of the questions investigated here represented a format used in a web survey prototype developed by the National Science Foundation in pursuit of changing a mail questionnaire to an electronic data collection instrument. The three questions chosen for experimentation had each presented problems to respondents in cognitive interviews of the prototype instrument.

The first issue investigated was whether the use of response category subheadings influenced respondent answers. Use of two headings which not only subcategorized response options topically, but resulted in graphical separation that made them appear as two distinct groups had a dramatic effect on people's answers. Many respondents felt compelled to provide answers in each sub-group when the headings and spacing were used (e.g. 70% marked an answer in both sub-groups, but only 41% did so when no headings or separation were used). We also found that the word instructions, "Please select the best answer," only partially mitigated the effect. Respondents checked significantly fewer items when the instructions were present (2.4 vs. 3.0), but were still just as likely to mark an answer in each of the sub-groups. It is obvious that visual grouping and separation have a dramatic effect on answers with or without the explicit instructions. At the same time, when the response choices were double banked on a different topic, very few differences were observed. These results suggest that separating a vertical list of categories into sub-groups is not a desirable construction practice. Doing so appears to set an expectation that the respondent provide an answer in each group. Thus, visual layout has an obvious and significant impact on responses.

The second issue focused on how the grouping of a large number of response options influences answering behavior. Of particular importance is the fact that the question asks for information that the respondent already knows, but must find among a long list of options. Asking for field of study or student major seems like a straightforward question, but the results indicate that how the options are grouped greatly influences the answering process. First, the results show that respondents can process some formats more efficiently than others. For example, a two-step grouping (organizing majors under general areas) is less efficient for respondents than a single-step grouping (alphabetical list) of the choices, but simply asking for an open-ended response and allowing respondents to report is the most efficient format. Second, differences in response patterns occur because of how response information is grouped. In the first survey, nearly 10 percent of respondents to the open ended question provided double or triple majors. When this was allowed in all versions of the second survey, slightly higher numbers of respondents indicated they had multiple majors (15% averaged across versions). The drop-down menu

used in the second survey provides visually separate places for listing first and second majors, and as such encouraged a greater number of people to do that (16.4%). Thus, we see time and response differences depending upon how the information is presented to respondents. It seems important to recognize that for what appears to be a straightforward question, response burden and response patterns differ greatly depending on how the question is visually presented to the respondent.

The third issue involved communicating to respondents the format they were expected to use when reporting a date. Results indicate that grouping the symbols with the answer spaces in order to communicate expectations is effective in conveying to respondents the desired format. Regardless of the position of the symbols, most respondents tended to use a two-digit month; however, when the symbols "MM YYYY" were placed directly below their respective answer boxes instead of to the right of the answer boxes respondents were significantly more likely to use a four-digit year (94% vs. 89.9%). This led to a larger proportion of respondents reporting the date using the desired format when the symbols were placed below the answer boxes (90.6 vs. 88.5) and overall, indicated that symbols are most efficient at communicating expectations when they are grouped, through proximity, with answer boxes.

In all three question types the grouping of information had significant effects on both the efficiency with which respondents were able to answer and the way they answered the questions. In addition, the evidence suggests that grouping interacts with other visual design elements as well as question formats and wording and as such its effect may vary depending on these other elements. Although much remains to be done to determine the effects of asking questions using different grouping schemes, the main conclusion to be reached at this point is that different approaches lead to different results and therefore need to be evaluated carefully before using them.

REFERENCES

Altheimer, Irshad and Don A. Dillman. 2002. "Results from Cognitive Interviews of NSF Earned Doctorate Web Survey." Social & Economic Sciences Research Center Tech Report #02-30.

Best, Samuel J. and Brian Krueger. 2003. *Internet Data Collection*. Unpublished Manuscript.

Christian, Leah Melani. 2003. "The Influence of Visual Layout on Scalar Questions in Web Surveys." Unpublished Master's Thesis, Washington State University.

Dillman, Don A. 2000. *Mail and Internet Surveys: The Tailored Design Method*. 2nd ed. New York, NY: John Wiley and Sons, Inc.

Dillman, Don A., Jolene D. Smyth, Leah M. Christian, and Michael J. Stern. 2003. "Multiple-Answer Questions in Self-Administered Surveys: The Use of Check-All-That-Apply and Forced-Choice Question Formats." Paper presented at the meetings of the American Statistical Association, San Francisco, CA, August 2003.

Heerwegh, D. & G. Loosveldt. 2002. "Describing Response Behavior in Websurveys Using Client Side Paradata." Paper Presented at the International Workshop on Web Surveys, Mannheim, Germany, October 2002.

Jenkins, Cleo R. and Don A. Dillman. 1997. "Towards a Theory of Self-Administered Questionnaire Design." in *Survey Measurement and Process Quality* edited by Lyberg, Biemer, Collins, de Leeuw, Dippo, Schwarz, and Trewin. John Wiley & Sons, Inc.

Neisser, Ulric. 1967. *Cognitive Psychology*. New York, NY: Meredith Publishing Company.

Palmer, Stephen E. 1999. *Vision Science Photons to Phenomenology*. Cambridge, MA: MIT Press.

Redline, Cleo D and Don A. Dillman. 2002. "The Influence of Alternative Visual Designs on Respondents' Performance with Branching Instructions in Self-Administered Questionnaires." in *Survey Nonresponse* edited by Robert M. Groves, Don A. Dillman, John L Eltinge, and Roderick J. A. Little. New York, NY: John Wiley & Sons.

Schwarz, Norbert. 1996. *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Schwarz, Norbert and Hans J. Hippler 1992. "Context Effects and Questionnaire Layout." Unpublished Research, ZUMA, Mannheim, Germany. Cited in *Cognition and Communication:Judgmental Biases, Research Methods, and the Logic of Conversation*. by Norbert Schwarz (1996).

Stern, Michael J, Don A. Dillman, Jolene D. Smyth, and Leah Melani Christian. 2004. "The Uses of Paradata for Evaluating Alternative Versions of Web Survey Questions." Paper presented at the 57th annual conference of the World Association for Public Opinion Research. Phoenix, Arizona. May 11-13, 2004