

An Examination of Visual Design Effects in a Self-Administered Mail Survey

Sarah Hastedt¹, Douglas Williams², and Catherine Billington, Ph.D.²

¹ National Center for Education Statistics, 1990 K Street NW, Washington, DC 20006

² Westat, 1600 Research Boulevard, Rockville, MD 20850

This paper is intended to promote the exchange of ideas among researchers and policy makers. The views expressed in this paper are part of ongoing research and analysis and do not necessarily reflect the position of the U.S. Department of Education.

Abstract

In self-administered surveys, respondents call upon a cognitive tool kit in terms of expectations about the question and answer process. In the context of mail surveys, employing design features that mimic these expectations (e.g., up is good) can work toward minimizing errors in response. This paper builds on previous research using data from the 2009 National Household Education Survey (NHES) Pilot Test which found variation in omission and commission error rates according to the visual design of the skip instructions on the self-administered mail questionnaire. In this paper, we look at experiments implemented in the design of skip instructions and the order of response categories using data from the 2011 NHES Field Test. The 2011 Field Test used a split-ballot questionnaire experiment which allows for comparisons of item-level nonresponse and response distributions across forms. In our first analysis, we examine whether increasing the emphasis of skip instructions has any effect on skip errors. In our second analysis, we examine the effectiveness of skip pattern design changes that eliminated the most problematic type of skip instruction used in the 2009 Pilot Test, a large highlighted box containing a skip instruction, from one of the questionnaire forms. In our third analysis we look at simple dichotomous responses to see if reversing the order affects distributions (e.g., 'yes/no' versus 'no/'yes'). Specifically, we examine whether switching the order of 'yes' or 'no' responses violates respondent expectations about what should come first and whether respondents rely more on their expectations than the survey when answering these questions.

Key words: questionnaire design, skip instructions, response option order

1. Introduction

Declines in landline telephone coverage rates and random digit dial (RDD) response rates in recent years have lead survey practitioners to consider alternate survey sampling frames and data collection modes, such as address-based sampling (ABS) with data collection via self-administered mail surveys. This trend underscores the importance of understanding how self-administered questionnaires can be designed to reduce respondent errors and improve the quality of data collected. Past research on questionnaire design has examined how a range of visual design features are related to survey response. For example, Christian and Dillman (2004) found that visual design variations such as the use of directional arrows, the size of write-in answer boxes, and the placement of instructions before or after response options influenced response. Research examining the effect of different skip instruction formats on respondent error rates has found that design characteristics of skip instructions can be manipulated to reduce error rates but that variation in error rates still exists across questions within a survey (Redline, Dillman, Dajani, and Scaggs 2003; Redline and Dillman 2002). In this paper we take

advantage of a split ballot questionnaire experiment implemented in the 2011 Field Test of the National Household Education Survey (NHES:2011 Field Test) to examine the effect of variations in the design of skip instructions and the order of response categories on navigation errors and response distributions.

1. 1 Overview of the NHES:2011 Field Test

The National Household Education Survey (NHES) is the primary household-based survey sponsored by the National Center for Education Statistics (NCES) within the U.S. Department of Education. It covers topics that are difficult to study in institution-based frames, such as early childhood care and education, children's readiness for school, parent involvement in education, and school choice. Surveys were conducted approximately every other year from 1991 through 2007. These prior administrations used random digit dial (RDD) sampling and telephone data collection from landline telephones. Declining response rates and concerns about population coverage with the landline telephone frame prompted NCES to redesign the NHES following the 2007 data collection. The NHES redesign involved shifting from an RDD landline sample with CATI data collection to an ABS frame with a two-phase mail survey as the primary data collection approach. This new methodology was first tested in a small-scale feasibility pilot test in 2009 (see Brick, Williams, and Montaquila (2011) for a detailed description).

The NHES:2011 Field Test was conducted to test strategies for improving response rates, refine operational procedures, and test different question wordings for certain items. The nationally representative sample comprised approximately 41,000 addresses¹. In the first phase of the Field Test, households were screened for the presence of eligible children. Sampled households were mailed an initial screener package which contained a cover letter, screener questionnaire, postage-paid return envelope, and an incentive. If there were no children age 20 or younger living in the household, respondents were asked to check a box indicating this and return the form. If there were children age 20 or younger living in the household, respondents were asked to enumerate each child and provide each child's age, sex, and school enrollment status. Households were sent up to two nonresponse follow-up mailings which contained a cover letter, replacement screener questionnaire, and postage-paid return envelope. In households whose returned screener questionnaires indicated the presence of at least one eligible child, one child was selected for a topical follow-up survey. Children ages 0 to 6 and not yet enrolled in kindergarten were eligible to receive the Early Childhood Program Participation (ECP) topical survey and children age 20 or under and enrolled in public or private school for kindergarten through twelfth grade were eligible for the Parent and Family Involvement in Education (PFI) topical survey. In households with multiple children, one child was randomly selected as the focal child for a topical survey in order to minimize burden on respondents. The initial topical mailing package contained a cover letter, topical questionnaire, and postage-paid return envelope. For some randomly selected respondents, the initial topical mailing contained a monetary incentive of varying amounts. Non-responding households were sent up to two topical follow-up mailings which included a cover letter, replacement topical questionnaire, and postage paid return envelope. For a subset of households that did not receive a monetary incentive with the initial mailing, the second follow-up mailing also contained an incentive.

¹ Two additional samples were drawn to examine aspects of the redesign associated with English literacy and the bilingual and Spanish materials. Analysis in this paper focuses only on the nationally representative sample.

1.2 Split-Ballot Experiment in the NHES:2011 Field Test

One of the experiments embedded in the NHES:2011 Field Test was a split-ballot test of two different questionnaire types for each topical survey (PFI and ECPP). Each topical form type had two versions, a ‘mainline’ version which primarily used the same question wording, skip pattern design, and response option order as the NHES:2009 Pilot Test and an ‘alternate’ version which used revised question wording for some items as well as different skip pattern designs and response option orders. Sampled cases were randomly assigned to receive either the mainline or alternate version of the questionnaire. While the split-ballot experiment was used to test different question wordings for some items, the analysis in this paper focuses on the influence of visual design aspects of the survey on response distributions and the following discussion will focus on differences between the mainline and alternate surveys in terms of their visual design.

One of the primary differences between the mainline and alternate questionnaires was the visual design of the skip instructions. Past research has consistently found that item nonresponse is higher on self-administered surveys that contain skip instructions compared to those that do not (Messmer and Seymour 1982, Turner et al. 1992, Gendall and Ramsay 2001). Respondents may make two types of errors due to skip instructions: commission errors and omission errors. Commission errors occur when respondents answer a question that they have been instructed to skip. Omission errors occur when respondents skip a question that they should have answered. The prevalence of problems with skip pattern compliance in self-administered surveys and the implications for data quality has motivated research that examines how different skip instruction formats are related to response error. Redline and Dillman (2002) tested the effect of three different types of skip instructions on item nonresponse in a classroom experiment setting. The control skip instruction used an arrow to the right of the response option pointing to the words “Skip to X” in italicized text. The second instruction tested, the detection branching instruction, was similar to the control instruction, but the skip instruction text was bolded and in larger font. An arrow was also added to the left of the non-branching response option to guide the respondent to the next question. The third skip instruction examined, the prevention branching instruction, used bold text for the verbal branching instruction against a white background. It also added instructions that reminded respondents to pay attention for branching instructions. Redline and Dillman (2002) found that both the detection instruction and the prevention instruction had a lower percentage of commission errors compared to the control design. In terms of omission errors, both the detection and prevention instruction designs had a slightly higher percentage of omission errors compared to the control instruction design. However, in a later experiment embedded in the 2000 Decennial Census of the United States which tested modifications to the instructions based on results of the classroom experiment, the detection instruction was found to lower both omission and commission errors (Redline et al. 2003).

In our research, the primary skip instruction in the mainline questionnaire consisted of an arrow to the left of the response option associated with the skip pointing to the text ‘GO TO question X’ (shown in figure 1²). The text was in the same size font as the response options, but was in bold and italicized font. An arrow to the left of the non-branching

² In some cases the question numbers differ across the mainline and alternate forms even though the item wording is the same. This due to the inclusion of new items in the alternate form for testing that were not included in the mainline form.

response option pointed to the next question. This design combined features of the control and prevention skip instruction designs tested in Redline and Dillman (2002). Combining features of these skip instructions was designed to take advantage of the benefits that each offered in terms of reducing respondent errors. Combining different skip instruction designs to further improve the performance of branching instructions was a direction for future research suggested in Redline et al. (2003). In the alternate form, a similar skip instruction was used with a notable revision intended to draw respondents' attention to the skip instruction text. As noted in Redline et al. (2003), in order for skip instruction text to be effective, respondents must first read the text. The revised skip instruction design used a colored box with a bold black outline to separate the skip instruction text from the question and response option text. The background color of the box containing the skip instruction text was dark green which contrasted with the lighter green background color of the questionnaire (figure 1 shows an example). The combination of the contrasting background color of the box and the bold black outline of the box was expected to draw respondents' attention to the box and the text contained within it. The expectation is that this change would reduce errors of omission and commission by drawing attention to the instruction and clarifying the navigational path. A potential drawback to this approach is the potential increase in item nonresponse to the gate question (containing the branching instruction) if respondents are drawn to the skip instruction rather than the item text and follow it without responding to the gate question. In the first analysis in this paper, we examine the effect of this revised skip instruction design on navigation errors.

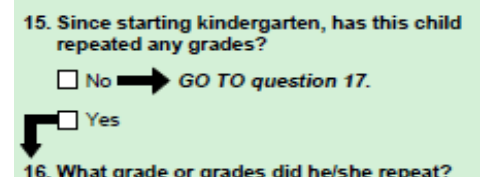
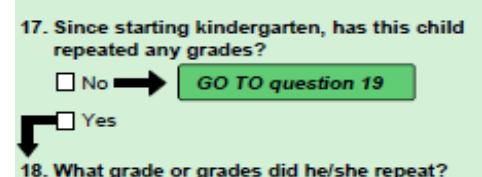
Mainline skip instruction design	Alternate skip instruction design
	

Figure 1: Skip instruction designs in the NHES:2011 Field Test

Some skip instructions are more complex and involve multiple dependencies or rely on more than the answer to a single item. Different versions of these more complex skip instructions were used in the mainline and alternate forms. The mainline versions used large highlighted stop boxes that were separate from the surrounding questions to direct respondents to skip over sections of the survey that did not apply to them. The stop box was a rectangle outlined with a black border with a yellow interior (shown figure 2). It contained bold text that told respondents how to proceed based on their response to the previous question. It also had an exclamation point to the left of the instruction to draw attention to it.

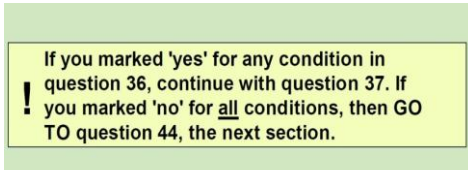
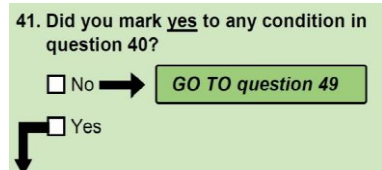
Mainline stop box design	Alternate skip design
	

Figure 2: Examples of stop boxes and embedded skips in the NHES:2011 Field Test

An analysis of data from an earlier pilot test of the new NHES methodology in 2009 found that the stop boxes used in the pilot test were associated with higher skip error rates than the other skip instructions used in the questionnaire (Redford and Hastedt 2011). In the alternate version of the NHES:2011 Field Test topical forms, these instructional stop boxes were converted to a question with an embedded skip instruction (shown in figure 2). In our second analysis we compare error rates between the two designs to examine whether the revised skip design reduced error rates relative to the stop box design.

Another difference between the mainline and alternate forms was the order of ‘yes’ and ‘no’ response options. In the NHES topical surveys, dichotomous items with ‘yes’ or ‘no’ response options generally asked about a behavior that the parent, the child, or the child’s school engaged in. The NHES topical surveys included opinion questions, but because these were generally Likert items they were not included in our analysis. In the mainline form, ‘yes’ was always presented before ‘no’ for items that were not a gate question for a skip. This followed the expected convention of ordering the positive response option first followed by the negative (Holbrook, Krosnick, Carson, and Mitchell 2000). The ‘yes’ and ‘no’ response options were reversed in some cases in the mainline when there was a skip and the ‘yes’ response continued to the next questionnaire item. This was done to visually connect the response to the next question the respondent should answer. However, when this order reversal is implemented intermittently, the unpredictable reversal of response options may cause confusion and increase the potential for error. In the alternate ‘no’ was always the first response followed by the ‘yes’ response to maintain consistency throughout the form.

Past research has found that when attitudinal response options were presented in an unconventional order respondents took longer to answer the questions (Holbrook et al. 2000). Holbrook and her colleagues also found that respondents gave answers that were less predictable and that were presumed to contain more error when the response options were presented in an unconventional order. Tourangeau, Couper, and Conrad (2004) found that response option orders which deviated from another convention that respondents use to navigate self-administered surveys, the ‘left and top mean first heuristic,’ were associated with slower response times in a Web survey. This heuristic captures the idea that if an item is the leftmost or top item in a list, it is conceptually the first item in the list and the subsequent options will follow in a logical order. Deviating from the ‘left and top means first’ heuristic was associated with a change in the response distribution, particularly for attitude items. For ‘yes’ and ‘no’ factual or biographical questionnaire items, there is generally a single answer that respondents are able to retrieve from memory. In many cases failure to retrieve an event from memory provides inference to the answer (Tourangeau et al. 2000). The factual nature of the items examined in this analysis suggests that the order of dichotomous response options may not impact response distributions in the context of the NHES:2011 Field Test. However, it is still important to understand the implications of using an unconventional response option order throughout the entire questionnaire for data quality. In this paper we looked at single item dichotomous response questions to examine whether switching the order of ‘yes’ and ‘no’ responses affected response distributions.

We also examined the impact of presenting the ‘no’ response before the ‘yes’ response in multiple part list-style items. Figure 3 shows an example of this type of question from the PFI alternate questionnaire. Response options for these questions were presented as two columns of answer boxes to the right of the question text with response category labels above each column. In the mainline the ‘yes’ column was first and the ‘no’ column was

to the right of it. The alternate versions of the topical surveys reversed this order with 'no' in the first column and 'yes' in the second column. This change was made to maintain consistency with single item questions where the first or top option was 'no.'

19. Has this child ever had the following experiences?		No	Yes
Mark <input checked="" type="checkbox"/> ONE box for each item below.		▼	▼
a. An out-of-school suspension.....	<input type="checkbox"/>	<input type="checkbox"/>	
b. An in-school suspension not counting detentions.....	<input type="checkbox"/>	<input type="checkbox"/>	
c. Been expelled from school.....	<input type="checkbox"/>	<input type="checkbox"/>	

Figure 3: Example of response option reversal in multiple-part items used in the PFI alternate survey of the NHES:2011 Field Test

In multiple part list-style questions that use a forced-choice format, questionnaire designers are often concerned that respondents will treat these questions as check-all items. This occurs when respondents only mark 'yes' responses and leave the 'no' category blank. In this case, item nonresponse may be an implicit 'no'; however, high levels of item nonresponse may reduce data quality and make the data difficult to interpret. Smyth, Dillman, Christian, and Stern (2005) found that approximately 2.7 percent of respondents treated forced choice items as check-all questions among the 24 questions they examined across two Web surveys and one paper survey. Presenting the response options in the unconventional 'no' then 'yes' order may reduce item nonresponse by increasing the salience of the 'no' option. Alternatively, respondents may overlook the response option order reversal and mark responses only in the first column which would impact response distributions. We looked at responses to multiple part list-style items to see if response distributions were affected and whether switching the order to make the 'no' response the first column increased its salience and reduced item nonresponse.

2. Methods and Data

The data used in this paper include all completed topical surveys that were part of the national sample from the NHES:2011 Field Test. The topical surveys included the PFI survey for children under age 20 and enrolled in kindergarten through twelfth grade and the ECPP survey for children ages 0 to 6 who had not yet begun kindergarten. For each survey, there was a mainline and alternate version. Sampled addresses were randomly assigned to each version. For the mainline there were a total of 5,150³ eligible⁴ households with an AAPOR RR1 of 74.0%. For the alternate there were a total of 2,250⁵ with an AAPOR RR1 of 73.5% (AAPOR 2011).

In order to ensure that there were not differences in respondent characteristics between the mainline and alternate samples of the ECPP and PFI that that could influence differences in error rates and response patterns independently from the design manipulations being tested, selected demographic characteristics were compared across

³ Sample size rounded to the nearest 50.

⁴ In order to be eligible for the topical phase a household must have responded to the screener survey and indicated that the household contained at least one child under age 6 who was not yet enrolled in kindergarten or at least one child under age 20 enrolled in school for grades K-12.

⁵ Sample size rounded to the nearest 50.

questionnaire versions within a particular topical type. Comparisons were only made within a topical type because there may be systematic differences in respondent characteristics related to the age of the children included in the sample for each topical type. Full results are available upon request. Overall, respondents to the mainline and alternate questionnaire versions of both the ECPP and PFI were very similar across the majority of characteristics examined. All analyses discussed in this paper used unweighted data, as we were not attempting to estimate error rates within the population. Thus the results presented here represent what we would expect for a similar sample population under similar conditions. Comparisons discussed in this paper were tested for statistical significance using two-tailed t-tests. Differences cited as statistically significant had a p -value less than .05. A list of items used in the analyses and the corresponding question text is available from the authors upon request.

3. Results

3.1 Effect of Increased Emphasis on the Skip Instruction

In our first analysis, we examined the effect of the revised skip instruction design in the alternate form on navigation errors. Table 1 presents omission, commission, and total error percentages by questionnaire version. The omission error percentage represents the percentage of respondents who made errors of omission given the opportunity to make this type of error. The commission error percentage represents the percentage of respondents who made errors of commission given the opportunity to make this type of error. The total error percentage is the percentage of respondents who made either an error of omission or an error of commission out of all respondents.

Table 1: Error Rates from Skip Emphasis Manipulation, by Type of Error and Questionnaire Version

Questionnaire Version	Total	Omission	Commission
Parent and Family Involvement in Education			
Mainline	5.0 (0.42)	2.3 (0.29)	8.5 (0.54)
Alternate	3.5* (0.54)	1.6 (0.37)	5.8* (0.69)
Early Childhood Program Participation			
Mainline	2.4 (0.47)	2.5 (0.48)	2.3 (0.46)
Alternate	1.7! (0.59)	1.1* (0.47)	2.7 (0.73)

* Significantly different from mainline version at $p < 0.05$.

! Interpret with caution. Standard error is more than 30 percent of estimate's value.

NOTE: Standard errors in parentheses.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2011 Field Test of the National Household Education Survey (NHES).

The error percentages in table 1 are based on an analysis of five items each in the PFI and ECPP. The items compared across versions had similar wording and placement in the questionnaire, but the visual design of the skip instruction varied across versions. On the PFI topical the total error percentage was lower on the alternate version than the mainline version (3.5% vs. 5.0%, respectively). For the ECPP topical the total error percentage was not significantly different across versions. For the PFI topical the commission error percentage was lower on the alternate questionnaire (5.8%) compared to the mainline (8.5%). On the ECPP, the only statistically significant difference in error percentages across questionnaire versions was in the omission error percentages; this difference was not significant for the PFI topical. The omission error percentage was lower on the ECPP alternate form (1.1%) compared to the mainline form (2.5%).

Our next concern was whether these changes affected item missing rates for the gate question (which included the skip instruction). The increased contrast between the skip box color and the background color may have drawn respondents' attention to the skip text itself rather than the question text, interrupting the response process for that item. For each item included in the first analysis, we looked at item nonresponse for the gate question (a total of five items in the PFI and five items for the ECPP). For the ECPP item nonresponse rates varied from 1.0% to 2.5% with an average the item missing rate of 1.9% in the mainline and 0.2% in the alternate for the questions considered. There were no significant differences in missing data rates for the questions examined in the ECPP. For the PFI, missing data rates across all selected items varied from 0 to 3.9%. On average, the item missing rate for the questions considered was 1.9% for the mainline and 1.2% for the alternate. The item missing rate differed significantly for one item on the PFI.

3.2 Effect of Change in Stop Box Design

Another design difference tested was the use of stop boxes in the mainline forms versus integrated skips in the alternate. Stop boxes are more complex skip instructions that are used to interrupt the navigational flow of the questionnaire and usually involve instructions that subset to a specific population group or require the respondent to consider two or more items to determine the appropriate path. In the mainline version of both the PFI and the ECPP, stop boxes were used to provide instructions to respondents regarding whether they should continue to the next question or skip to another section. They were designed to draw attention to the skip instruction by the use of color, font, and symbols that provided a contrast between the box and the surrounding questionnaire items. Previous research suggested that errors were greater with stop box skip instructions compared to skip instructions embedded in the question (Redford and Hastedt 2011). In the alternate versions of the PFI and ECPP, the stop boxes were replaced with a question with an embedded skip, similar to that used throughout the rest of the questionnaire.

There were only two questions that used the stop box design in both the PFI and the ECPP. The results presented in table 2 are based on these two questions. The mainline error percentages are for the questions immediately following the stop box and the alternate version error percentages are the error percentages for the follow-up question to the item used in place of the stop box. For both the PFI and the ECPP alternate forms errors of omission were largely eliminated. For the PFI the omission error rate was 0.8% compared to 5.9% in the mainline version, and for the ECPP omission errors were nonexistent (no errors at all compared to 5.2% for the mainline version). Errors of

commission were much higher in the alternate forms than in the mainline forms for both topical types (24.7% versus 13.1% for the PFI and 28.3% versus 15.7% for the ECPP).

Table 2: Error Rates from the Stop Box Manipulation, by Type of Error and Questionnaire Version

Questionnaire Version	Total	Omission	Commission
Parent and Family Involvement in Education			
Mainline	11.9 (0.62)	5.9 (0.45)	13.1 (0.65)
Alternate	20.4* (1.19)	0.8*! (0.26)	24.7* (1.28)
Early Childhood Program Participation			
Mainline	14.4 (1.08)	5.2 (0.68)	15.7 (1.11)
Alternate	24.8* (1.95)	--	28.3* (2.04)

* Significantly different from mainline version at $p < 0.05$.

! Interpret with caution. Standard error is more than 30 percent of estimate's value.

NOTE: Standard errors in parentheses.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2011 Field Test of the National Household Education Survey (NHES).

Because of the high commission error percentages observed on the alternate forms, we examined the error rates at the item-level. The first item we examined asked if the sampled child was receiving services for any of the health conditions listed in the previous question. The commission error rate for the follow-up question to this item was 11.5% on the PFI mainline and 17.9% on the ECPP mainline. These error rates were much higher than those observed on the alternate version of the same item in the PFI (3.1%) and ECPP (3.4%). In the second question examined, respondents were asked if the sampled child was enrolled in any programs for English language learners. On the follow-up item for this question, the commission error rate was much higher on the PFI and ECPP alternate forms (41.2% and 52.4%, respectively) compared to the PFI and ECPP mainline versions (14.0% and 12.9%, respectively).

The difference in commission error rates between these questions suggests that respondents use more than design features, such as skip instructions, in determining whether a question is appropriate. We cannot determine the reason for the difference in results with this analysis, but we can speculate that the differences observed may be in part due to characteristics of the follow-up questions. One question examined asked whether the child is receiving services for his or her health condition. If respondents who are instructed to skip this item fail to follow that instruction, they may find it difficult to answer this item as the response that best fits their child is not available. There is an implied cue to the respondent that this question does not apply. In the other item examined, respondents were asked if the child was enrolled in bilingual or ESL programs. If respondents fail to follow the skip, an appropriate response is available that

does not contradict any previous response. However, Redline et al. (2003) found that respondents answered questions that did not apply to them despite the cues contained in follow-up questions suggesting that differences in cues between questions alone may not explain our findings. Another potential explanation for the conflicting results is related to differences in the cognitive burden of the skip instruction used to replace the stop box. The item with lower commission error percentage was converted to a dichotomous response question that used the same revised skip instruction design as the rest of the alternate questionnaire. The item with the higher commission error rate replaced the stop box with a check-one response list. There was a longer list of response options in this question and the options themselves contained more text. The skip instruction used the same highlighted box and black outline design as the other skip instructions, but rather than having an arrow point to the skip instruction text, a bracket to the right of two response options pointed to the skip instruction text. This may have been more cognitively demanding for respondents to follow.

3.3 Effect of reversing the ‘yes’/‘no’ response order

In order to examine the effect of deviating from the expected ‘yes’/ ‘no’ response order on the response distribution, we compared the percentage of ‘yes’ responses to selected items on the mainline and alternate forms. A total of 9 items were selected from the PFI form and a total of 8 items were selected from the ECPP form. The items selected covered a diverse set of topics, were not embedded within a skip that may confound findings, and included similar wording between versions. In the mainline forms, ‘yes’ was generally the first or top response, with ‘no’ as the second or bottom response. In the alternate form ‘no’ was always the first or top option and ‘yes’ was the second or bottom option. The motivation for this change was to increase the visual connection between non-branching response options (usually ‘yes’) and the next applicable question for items that were part of a skip pattern. To maintain consistency throughout the form the order of response categories was reversed for all items in the alternate. A concern with this is that if respondents have an expectation that the ‘yes’ response should be first, this change in order will affect response distributions. While all response options were labeled in the questionnaires, respondents may overlook these and infer meaning from the position of the response options.

Table 3 presents the percentage of ‘yes’ responses for selected items on the PFI. When we examined the response distributions for the ECPP form, there were no statistically significant differences between the percentage of ‘yes’ responses on the ECPP mainline and alternate versions for the questions examined in this analysis. Due to space constraints these results are not shown here but are available upon request. On the PFI only one item examined showed a statistically significant difference in the percentage of ‘yes’ responses between the alternate and the mainline. This item asked if the respondent considered other schools for the sampled child. The percentage of ‘yes’ responses on the mainline version was 35.1% compared to 19.2% on the alternate. This is a rather large difference and conflicts with the lack of differences found for the other items. We hypothesize that this difference is not due to differences in response option order, but rather a context effect because the alternate version of the PFI contained a new question before this item that was not in the mainline version. This new question asked whether or not the respondent’s public school district allowed parents to choose which public school to send the child to. This question may have created a context effect by influencing the way that respondents defined the concept of considering other schools for the child. If respondents defined considering other schools more narrowly as a result of this question,

this could explain the lower percentage of respondents reporting that they considered other schools for the child on the PFI alternate.

Table 3: Distribution of ‘Yes’ Responses from the Parent and Family Involvement in Education Survey, by Questionnaire Version and Selected Items

Questionnaire Item	Percent ‘Yes’	
	Mainline	Alternate
Did you move to your current neighborhood so your child could attend his/her current school?	19.0 (0.77)	19.0 (1.17)
Did you consider other schools for this child?	35.1 (0.93)	19.2* (1.18)
Since the beginning of this school year, has this child been in the same school?	97.1 (0.33)	96.5 (0.55)
Since starting kindergarten, has this child repeated any grades? ¹	8.1 (0.53)	8.2 (0.82)
Is this child of Spanish, Hispanic, or Latino origin?	16.9 (0.73)	15.9 (1.10)
Other than this address, does anyone in this household currently receive mail at another address including P.O. Boxes?	13.6 (0.66)	11.7 (1.10)
Do you have access to the Internet at this address?	89.6 (0.59)	88.5 (.95)
Is there at least one telephone in this household that is currently working and not a cell phone?	76.9 (0.82)	74.2 (1.30)
Do you have a working cell phone?	94.9 (0.43)	94.5 (0.68)

¹ Response order of ‘yes’/‘no’ is the same between mainline and alternate versions

* Significantly different from mainline version at $p < 0.05$.

NOTE: Standard errors in parentheses.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2011 Field Test of the National Household Education Survey (NHES).

Our final analysis compared the distribution of ‘yes’ responses for multiple-part items presented as a list. Table 4 shows the distribution of ‘yes’ responses for selected items for the PFI forms. We do not present results for the ECPP forms because items of this format were included in sections that only a small number of respondents were asked. Only items with a minimum set of three that were not embedded within a skip instruction were included in the analysis. Due to space constraints, results for seven of the fourteen items examined are shown in table 4. The full results are available upon request.

Significant differences were observed in four of the fourteen questions, with at least one significant difference from each set. Three of the four significant differences were for questions near the end of the list. The lack of significant differences between items early

in the list suggests that respondents are noting and adhering to the reversal of the response options; however, differences toward the end of the set suggest that respondents may be reverting to order expectations as the response labels become visually more distant.

Table 4: Distribution of ‘Yes’ Responses and Percent Missing for the Parent and Family Involvement in Education Survey, by Questionnaire Version and Selected Item Sets

Questionnaire Item (Set)	Percent ‘Yes’		Percent missing	
	Mainline	Alternate	Mainline	Alternate
Set 1				
Has child had out-of-school suspension?	7.3 (0.51)	8.6 (0.84)	2.0 (0.27)	1.5 (0.36)
Has child had in-school suspension?	8.7 (0.55)	9.6 (0.88)	2.1 (0.28)	2.3 (0.44)
Has child been expelled from school?	0.8 (0.17)	2.7* (0.49)	3.0 (0.33)	2.5 (0.46)
Set 2				
Has anyone worked on a project with child?	54.0 (0.98)	54.6 (1.50)	4.0 (0.38)	2.8 (0.49)
Has anyone played sports with child?	71.4 (0.89)	72.6 (1.33)	3.8 (0.37)	1.6* (0.37)
Has anyone discussed time management with child?	68.4 (0.91)	69.0 (1.38)	3.8 (0.37)	1.5* (0.36)
Has anyone talked about the family’s history with child?	47.2 (0.98)	53.5* (1.49)	3.6 (0.36)	1.7* (0.38)

* Significantly different from mainline version at $p < 0.05$.

NOTE: Standard errors in parentheses.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2011 Field Test of the National Household Education Survey (NHES).

We next looked at how item nonresponse was affected by the reversal of responses for dichotomous items presented as a list. These items used a forced-choice format which has been observed to have substantial rates of item nonresponse. Respondents may treat the forced choice list as mark-all-that-apply, thus missing responses may be implicit ‘no’ responses. In the alternate version of the PFI questionnaires, presenting ‘no’ in the first column may have affected the saliency of the ‘no’ response thereby increasing the perception that a response is needed for ‘no.’ If this is the case we would expect item nonresponse to be reduced in the alternate version relative to the mainline version. Differences in item nonresponse were only significant for three of the 14 items, and all three items were included within one question set. For this set of items there was a stop box at the beginning of this section for a previous set of items in the mainline form applicable to parents of elementary school children. It is possible that this skip was not followed correctly thereby increasing item nonresponse for the list included in our analysis. Excluding this list and focusing on the remaining nine items, none of the differences are significant. However, for eight of the nine items, the difference is in the

expected direction of reduced item nonresponse. While we cannot conclude with these findings that reversing the order of 'yes'/'no' responses for multiple item lists reduces item nonresponse, these results suggest that possibility.

4. Discussion

As declining response rates to telephone surveys lead to shifts to other modes of data collection, it is important to understand how visual design aspects of questionnaires can be optimized to minimize respondent errors and obtain high quality data. The split-ballot experiments used in the NHES:2011 Field Test provided an opportunity to examine the effect of several visual design variations on navigation errors, response distributions, and item nonresponse in a self-administered mail survey.

The first design features we focused on in our analysis were changes to the skip instruction design that increased the contrast between the skip instruction box color and questionnaire background color. These changes were intended to draw attention to the skip instruction and reduce errors of omission and commission. A potential drawback of increasing the focus on the skip instruction is an increase in item nonresponse in the gate question if respondents follow the skip without first responding to the gate question. On the PFI topical, the commission error percentage was approximately 30% lower on the alternate questionnaire compared to the mainline. On the ECPP, the omission error percentage was more than 40% lower on the alternate form compared to the mainline form. The finding that the omission error percentage was lower on the ECPP alternate form compared to the mainline is promising, as omission errors are generally believed to be the more problematic type of error because they result in missing data. For the gate questions examined in our analysis item nonresponse rates were generally low and did not significantly differ across the mainline and alternate forms with the exception of one item.

Color is often used in surveys to indicate areas within the questionnaire that are relevant to the question and response process. The difficulty with the use of colors in surveys is to find balance between subtlety and saliency. In this experiment the slightly more contrasting color used in the skip instruction box compared to the questionnaire background to emphasize the navigational instructions was associated with reduced errors of commission on the PFI alternate form and reduced errors of omission on the ECPP alternate form. The logic of this finding makes sense in the context of errors of commission. The increased saliency draws respondents' attention and respondents correctly go to the next appropriate question. However, for errors of omission respondents are choosing to skip appropriate questions. For the items we examined errors of omission are the result of navigating beyond the next question. Increasing the salience of the skip instruction could have given the expectation of increasing errors of omission, but we observed either significant decreases or nominal decreases. One theory is that the increased saliency caused respondents to double-check whether they should follow the skip or not. Another potential explanation is that the increased saliency made the connection between the skip option and skip instruction more well-defined, thus clarifying for respondents that they are not supposed to skip.

The second design feature we examined was the elimination of stop box skip instructions in the alternate form. For both the PFI and the ECPP alternate forms, errors of omission were largely eliminated by replacing the stop boxes with an integrated skip instruction. However, errors of commission were higher in the alternate forms than in the mainline forms for both the PFI and ECPP. An item-level analysis revealed that the increase in commission errors was accounted for by one question. For the other question included in

the analysis, errors of commission were significantly lower in the alternate form. Differences in the cognitive burden of the skip or the saliency of the skip used to replace the stop box may explain the conflicting results observed for errors of commission. The current analysis cannot disentangle the effect of differences in question complexity and differences in the complexity of the skip instruction itself; more research is needed to disentangle the reason for the conflicting results found here. It is important for questionnaire designers to consider the context in which particular items are asked and how this may interact with visual design features of the questionnaire in order to minimize respondent error and collect high quality data. This is in line with the findings of Stern, Smyth, and Mendez (2011) who found that question saliency has response effects that are independent of the question layout and effects that are related to the question layout.

The third design feature we examined was the reversal of dichotomous response options. The NHES topical surveys contained a large number of dichotomous response items, many of which were associated with a skip instruction due to the length and complexity of the questionnaire. Reversing the conventional 'yes' then 'no' response option order was intended to increase the visual connection between non-branching response option (generally 'yes') and the subsequent questionnaire item. For the single question dichotomous response items we examined, we did not see evidence that reversing the order of 'yes'/'no' responses affected the response distributions. This may be because the items we examined asked about factual constructs that were related to behaviors or attributes that were relatively easy for respondents to recall rather than attitudes which may be more affected by changes in response options (Tourangeau et al. 2004). For the 'yes'/'no' list format questions, the majority of the questions examined did not have a significantly different percentage of 'yes' responses across the mainline and alternate forms. However, of the significant differences observed, the majority were for questions at or near the end of the list of questions in a particular set. Overall, the findings on the reversal of the 'yes'/'no' response options suggest that for single dichotomous response questions about factual constructs, deviating from the expected response order does not affect response distributions. However, for list-style response options, deviating from the expected 'yes'/'no' order may have response distribution implications for items near the end of the list.

Overall, we found that the revised skip instruction design used in the alternate form of the NHES:2011 Field Test was effective at reducing total and commission error rates on the PFI survey and omission error rates on the ECPP survey compared to the design used in the mainline forms. We found mixed results regarding the effect of revisions to the stop box skip instruction design. Our findings suggest that the cognitive burden and saliency of the skip may interact with the question content to influence response. Response option order changes made to questions with dichotomous responses generally did not result in significant differences in response distributions. However, questionnaire designers should exercise caution when implementing response option orders that run counter to respondent expectations. The findings for list-style items with dichotomous responses are suggestive that respondents adhere to response indicators within the questionnaire, but when these indicators are visually distant respondents may fall back on heuristic expectations. Additional research is needed to understand how visual design features interact with other survey features to improve questionnaire design and collect higher quality data.

References

- The American Association for Public Opinion Research. 2011. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 7th edition. AAPOR.
- Brick, J.M., Williams, D., and Montaquila, J.M. (2011) Address-Based Sampling for Subpopulation Surveys. *Public Opinion Quarterly*, 75(3): 409-428.
- Christian, L.M., and Dillman, D.A. (2004). "The Influence of Graphical and Symbolic Language Manipulations on Responses to Self-Administered Questions." *Public Opinion Quarterly* 68:57-80.
- Gendall, P. and Ramsay, R. (2001). An Experiment with Skip Instructions, *Marketing Bulletin*, 12.
- Gentry, R., Cantave, M., Wasikowski, N., and Pens, Y. (2010). "To Mail or to Call: How to Reach the Hard-to-Reach." Paper presented at the 65th Annual Meeting of the American Association for Public Opinion Research, Chicago.
- Holbrook, A.L. Krosnick, J.A., Carson, R.T., and Mitchell, R.C. (2000). Violating Conversational Conventions Disrupts Cognitive Processing of Attitude Questions. *Journal of Personality and Social Psychology*, 36: 465-494.
- Iannacchione, V.G. (2011). The Changing Role of Address-Based Sampling in Survey Research. *Public Opinion Quarterly*, 75(3): 556-575.
- Link, M.W., Daily, G., Shuttles, C.D., Yancey, L.T., Burks, A.T., and H. Christine Bourquin, H.C. (2009). "Building a New Foundation: Transitioning to Address-Based Sampling after 30 years of RDD." Paper presented at the 64th Annual Meeting of the American Association for Public Opinion Research, Hollywood, FL.
- Messmer, D.J., and Seymour, D.T. (1982). The Effects of Branching on Item Nonresponse. *Public Opinion Quarterly*, 46(2): 270-277.
- Redford, J. and Hastedt, S. (2011). Who Missed the Skips?: Empirical Results from a Self-Administered Survey. Paper presented at the 66th Annual Meeting of the American Association for Public Opinion Research, Phoenix, AZ.
- Redline, C.D. and Dillman, D.A. The Influence of Alternative Visual Design on Respondents' Performance with Branching Instructions in Self-Administered Questionnaires. In: Groves R., Dillman D., Litte R., Eltinge J., editors. *Survey Nonresponse*. New York: John Wiley; 2002.
- Smyth, J.D., Dillman, D.A., Christian, L.M., and Stern, M.J. (2005) Comparing check-all and forced-choice question formats in web surveys. *Public Opinion Quarterly*, 70(1): 66-77.
- Stern, M.J., Smyth, J.D., and Mendez, J. (2012). The Effects of Item Saliency and Question Design on Measurement Error in a Self-Administered Survey. *Field Methods*, 24(1): 3-27.
- Tourangeau, R., Couper, M.P., and Conrad, F. (2004). Spacing, Position, and Order. Interpretive Heuristics for Visual Features of Survey Questions. *Public Opinion Quarterly*, 68(3): 368-393.
- Tourangeau, R., Rips, L.J., & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Turner, C. F., Lessler, J. T. & Devore, J. (1992) Effects of mode of administration and wording on reporting of drug use. In: Turner, C. F., Lessler, J. T. & Gfroerer, J. C., eds. *Survey Measurement of Drug Use: Methodological Issues*, pp. 177-219. DHHS Publication Number, pp. 92-1929. Washington, DC: Government Printing Office.