Professionally Designed Questionnaires: Do they collect better data?

Kathy Ott and Chadd Crouse, National Agricultural Statistics Service
3251 Old Lee Highway, Room 305, Fairfax, VA 22030
kott@nass.usda.gov, (703) 877-8000, x117

Abstract

The National Agricultural Statistics Service (NASS) collects economic data from farm operations throughout the country on an annual, detailed, lengthy survey called the Agricultural Resource Management Survey. The data include extensive information on farm debt, assets, expenses, and income, as well as more general personal debt and income information. Traditionally, this survey has been conducted using personal interviewers who made face-to-face contacts to gather the data on a 30+ page questionnaire.

As part of a mode experiment, in 2002, a small amount of data for this survey were collected through the mail using a shortened version (8 pages) of the interviewer administered form, that was not specifically designed for use as a self administered form. In 2004, the need for larger samples in certain states coupled with a limited budget for the survey made it necessary to use mail data collection on a large scale.

A new, professionally designed, 16 page questionnaire was developed by the Social and Economic Sciences Research Center at Washington State University for use as both a mail and face-to-face instrument for this data collection. This instrument incorporated a variety of visual design concepts and innovative formatting for particular questions.

Item nonresponse and estimates for particular variables from the original 2002 data collection will be compared to data from the newly designed questionnaire to determine if the visual design and formatting concepts were effective at obtaining more accurate data.

Keywords: Visual design concepts, mode effects, self administered surveys, economic data

1. Introduction

Questionnaire design is an important part of data collection for surveys. The design of the survey instrument can theoretically influence whether respondents will fill out the questionnaire, whether they get frustrated while filling it out, and the quality of the data they provide. This analysis compares results from two questionnaire versions to determine if a comprehensive re-design impacted data quality for the survey.

The National Agricultural Statistics Service (NASS) is the statistical data collection agency of the United States Department of Agriculture (USDA). NASS conducts over 400 surveys annually covering all facets of US agriculture including acreage and production, livestock inventory, and farm economics. One of the surveys conducted each year is Phase 3 of the Agricultural Resource Management Survey (ARMS), a joint effort with the Economic Research Service, another USDA agency. This survey collects detailed information on farm and ranch economics including debt, assets, income, expenses, household characteristics and operator demographics. From this phase of the ARMS survey, regional and select state level estimates are published.

Traditionally, a 'long form' questionnaire is used for ARMS data collection. This form is typically 32-36 pages long and is used in face-to-face interviews that last an average of 90 minutes. In order to increase the sample size while minimizing the increase in respondent burden and still containing costs and maintaining steady interviewer workloads, a self-administered questionnaire was developed and tested in 2001. This eight page 'short form' questionnaire was used in conjunction with a four-page instruction booklet in a mailout/mailback data collection methodology. Basically, the short form was created by combining categories from the long form and eliminating questions that vary from year to year. For example, the long form questionnaire asks operators to report their fuel expense in several categories including diesel, oil, gasoline, natural gas, and LP gas. The short form asked one question that included all those types of fuel in the same answer.

The short form questionnaire was used in five states to test mailout/mailback data collection methodology for the ARMS survey.

In 2003, the original short form self-administered questionnaire was re-designed by Dr. Don Dillman and Dr. Danna Moore of the Social and Economic Sciences Research Center at Washington State University. The goals of the redesign were to:

- keep the content the same,
- make the questionnaire more user-friendly,
- incorporate the instruction booklet into the questionnaire, and
- standardize the formatting.

In addition, NASS imposed a 16-page limit on the length of the new questionnaire. This was an agency decision based on the belief that a longer self-administered questionnaire would be detrimental to respondent perceptions of burden. As stated earlier, the 2001 questionnaire was eight pages, but included a four-page instruction booklet that had to be referenced in order to complete the survey.

For some examples of how the questionnaire was changed, see Appendix A. For the theoretical basis of the changes, refer to *Achieving Usability in Establishment Surveys Through the Application of Visual Design Principles* by Dillman, Gertseva and Mahon-Haft (2005).

The redesigned questionnaire was used in 15 states, two of which (Minnesota and Texas) overlapped with the states involved in the 2001 test. These are the only two states included in the analysis done for this study.

2. Experiment

The overall goal was to compare the 2001 Short Form to the 2003 Short Form. The two questionnaires were not directly compared because agricultural data change greatly over time, even over one year. Data from surveys two years apart could easily be different just due to changes in agriculture, weather, or farm policies. Therefore, for each questionnaire comparison, the long form versions of the questionnaire for each year were used as a baseline and were compared to the short form version for that same year.

To do the comparisons then, the 2001 Short Form was compared to the 2001 Long Form and the 2003 Short Form was compared to the 2003 Long Form. Two data quality tests were performed for each year. One compared the distribution of data for 15 variables and the other compared item nonresponse for nine variables.

Before discussing the data quality tests that were completed, there are several items of caution that should be mentioned. First, this study was not set up as a controlled experiment. Data were collected in 2001 with the 2001 short form, and independently in 2003 using the 2003 short form. Second, there were methodological differences across the different years and different questionnaires. Both the 2001 and 2003 long forms were administered exclusively using face-to-face interviewing. The 2001 Short Form was administered through the mail with two mail contacts. The 2003 Short Form was administered with a more comprehensive mail methodology as well as face-to-face follow-up interviews. Third, the data were keyed after they were partially edited by NASS staff, and therefore, the data on our analysis file were not necessarily the same as those originally reported. Next, not all variables were on all four versions of the ARMS questionnaires. Only variables on all versions could be analyzed for this study. Finally, there are only a small number of records for analysis in some cases because only two states participated in both data collection years of interest – Minnesota and Texas. The 2001 short form sample sizes were 75 records in Minnesota and 150 records in Texas. The sample sizes in 2003 were 1,022 in Minnesota and 956 in Texas. The sample size for 2003 was much larger than that in 2001 because the 2001 sample was a test, while the 2003 sample was used operationally for production data.

3. Data Quality Tests

In order to determine whether one questionnaire performed better than another, two different data quality tests were used. Statistical testing was done to determine if there were differences in the distribution of data and item non-response rates for the short form and long form for each year. If one questionnaire had more statistical differences than the other for one of the tests, we may be able to conclude that the form is not as effective a data collection instrument as the other questionnaire.

3.1 Distribution of Data

The first data quality test performed was used to determine if the data from each questionnaire version are distributed similarly. A standard t-test could not be used because the data for most ARMS variables are not normally distributed. In fact, many are clustered at zero and include large outliers. Therefore, the non-parametric, distribution-free Kruskal-Wallis Test was used. To perform the Kruskal-Wallis Test, the samples to be compared are combined and all observations in the set are ranked. The test statistic is a function of the average rank in the combined set of observations from each sample. If data from each sample have the same distribution, ranks will be approximately the same.

Fifteen variable distributions were compared. These fifteen variables were selected because they were available on all four questionnaires without doing complicated calculations. These variables can be divided into the following six categories:

-Demographics:
    -Age
    -Education
    -Year started operating
-Acreage:
    -Total Acres
-Assets:
    -Total assets
    -Estimated market value (EMV) of production inputs – beginning of year
    -EMV of production inputs – end of year
    -EMV of trucks and cars owned
-Income:
    -Total Income
    -Other income (a catch-all question for any income not reported anywhere else)
    -Rent received
    -Government payments (for agriculture related programs)
-Debt:
    -Total debt
-Expenses:
    -Operator expenses
    -Cash rent paid

Minnesota and Texas were analyzed separately, so a total of 30 comparisons were done. No adjustments we made for the number of tests done. We considered adjusting for the large number of tests. However, due to the small number of significant tests, the adjustment would not have resulted in any practical differences in our results.

We assumed that the long form values for each variable for each year were the truth. Therefore, if one short form questionnaire had a larger number of variable values that were significantly different from the long form for a particular year than the other short form questionnaire, its data would be considered worse.

There were few significant differences at the $\alpha < 0.05$ level for either questionnaire. There were four in 2001 and six in 2003, two of which overlapped. The specific variables for each year are shown in Tables 1 and 2 (all tables are in Section 5 of this paper). An '*' in the last column indicates those variables where $\alpha < 0.05$.

3.2 Item Nonresponse

The second data quality test done was Fisher's Exact Test to determine whether the item nonresponse for specific variables was different for the Short Form and Long Form for a certain year. Six variables were analyzed for this data quality test. These were a subset of the 15 variables used for the distribution of data test. The remaining nine variables are those that have no item nonresponse because NASS requires a value to be present for all records. If a value is not present when the form is received, a staff member will make a follow-up phone call or use other information in the questionnaire to enter a reasonable number. The six variables used for nonresponse analysis were:

-Assets
    -EMV of production inputs – beginning of the year
    -EMV of production inputs – end of the year
    -EMV of trucks and cars owned
-Income:
    -Other income
    -Rent received
-Expenses:
    -Cash rent paid

For each of the four questionnaires, the proportion of zero values was calculated for each variable. Then, pair-wise comparisons were done between:

- the 2001 Short Form and the 2001 Long Form, and
- the 2003 Short Form and the 2003 Long Form.

Keep in mind that this test just indicates a difference, but does not indicate which questionnaire is better. Also, as mentioned earlier, we did not have data exactly as it came back from the respondent. NASS staff often check through questionnaires before they are keyed to be sure that all crucial items are filled in and that other values are reasonable. If certain variables are not filled in or are inconsistent, the office staff try to determine a sensible value.

There were few significant differences ($\alpha < 0.05$) between the item nonresponse rates for the questionnaires. There were three differences for 2001 and four for 2003, two of which overlapped. The specific variables and their values for each year are shown in Tables 3 and 4 (all tables are in Section 5 of this paper). The percentages show the missing or *nonresponse* rates for each variable.

The two short form questionnaires have a similar number of differences at the $\alpha < 0.05$ level; the 2001 questionnaire had 4 differences, the 2003 had 3 differences. Since both questionnaires have about the same small number of differences, we cannot conclude that one is better or worse than the other for general item nonresponse.

4. Conclusion

It was difficult to compare the two questionnaire versions because of several factors including sample sizes, methodologies used and the availability of data from each survey. Still, the comparison was done in an attempt to quantify any data changes resulting from the questionnaire redesign.

Based on the results of the data quality tests performed for these two questionnaires, it is inconclusive which questionnaire is better. There are differences in the data distributions for some items, but the differences are not concentrated in one questionnaire or the other. In fact, they have an almost identical number of differences between the short and long form for each year.

Although neither questionnaire performed better or worse in the data quality tests that were used, from a questionnaire design perspective, the redesigned form is much more visually appealing and is user-friendly as a self-administered instrument. The redesigned form incorporates a wide variety of design elements that make it easier to navigate and fill out for respondents. For these reasons, the redesigned form will be used for future surveys and will be

used as a guide for re-designing other versions of questionnaires used at NASS.

5. Reference

Dillman, Don A., Arina Gertseva and Taj Mahon-Haft. *Achieving Usability in Establishment Surveys Through the Application of Visual Design Principles*. Journal of Official Statistics, Vol. 21, No. 2, 2005, pp. 183-214.

6. Tables

Table 1: Distribution comparisons for 2001

State/Variable	p value	
MN 'Assets'	p=0.1748	
MN 'Operator Expenses'	p=0.0864	
MN 'Total Acres'	p=0.0671	
MN 'Cash Rent Paid'	p=0.2054	
MN 'Other income'	p=0.0000	*
MN 'Rent Received'	p=0.2876	
MN 'EMV of production inputs – beginning of the year'	p=0.9640	
MN 'EMV of production inputs – end of the year'	p=0.8882	
MN 'EMV of trucks and cars'	p=0.3385	
MN 'Age'	p=0.6015	
MN 'Education level'	p=0.3312	
MN 'Year Started Operating'	p=0.5229	
MN 'Income'	p=0.0752	
MN 'Debt'	p=0.7525	
MN 'Government Payments'	p=0.0454	*
TX 'Assets'	p=0.4617	
TX 'Operator Expenses'	p=0.1026	
TX 'Total Acres'	p=0.8450	
TX 'Cash Rent Paid'	p=0.8996	
TX 'Other income'	p=0.0000	*
TX 'Rent Received'	p=0.9251	
TX 'EMV of production inputs – beginning of the year'	p=0.1165	
TX 'EMV of production inputs – end of the year'	p=0.2081	
TX 'EMV of trucks and cars'	p=0.0545	
TX 'Age'	p=0.7718	
TX 'Education level'	p=0.3157	
TX 'Year Started Operating'	p=0.6995	
TX 'Income'	p=0.0147	*
TX 'Debt'	p=0.3515	
TX 'Government Payments'	p=0.7008	

* significant that the $\alpha < 0.05$ level

Table 2: Distribution comparisons for 2003

State/Variable	p value	
MN 'Assets'	p=0.2808	
MN 'Operator Expenses'	p=0.8632	
MN 'Total Acres'	p=0.3432	
MN 'Cash Rent Paid'	p=0.4556	
MN 'Other income'	p=0.0001	*
MN 'Rent Received'	p=0.3039	
MN 'EMV of production inputs – beginning of the year'	p=0.0267	*
MN 'EMV of production inputs – end of the year'	p=0.0052	*
MN 'EMV of trucks and cars'	p=0.4600	
MN 'Age'	p=0.1088	
MN 'Education level'	p=0.5553	
MN 'Year Started Operating'	p=0.1235	
MN 'Income'	p=0.7320	
MN 'Debt'	p=0.6603	
MN 'Government Payments'	p=0.6178	
TX 'Assets'	p=0.7528	
TX 'Operator Expenses'	p=0.0086	*
TX 'Total Acres'	p=0.1333	
TX 'Cash Rent Paid'	p=0.6709	
TX 'Other income'	p=0.4546	
TX 'Rent Received'	p=0.1023	
TX 'EMV of production inputs – beginning of the year'	p=0.6632	
TX 'EMV of production inputs – end of the year'	p=0.1726	
TX 'EMV of trucks and cars'	p=0.8766	
TX 'Age'	p=0.2610	
TX 'Education level'	p=0.9720	
TX 'Year Started Operating'	p=0.7814	
TX 'Income'	p=0.0454	*
TX 'Debt'	p=0.8907	
TX 'Government Payments'	p=0.0149	*

* significant that the $\alpha < 0.05$ level

Table 3: Percentage of item nonresponse for 2001

State/Variable	Short Form %	Long Form %	p value	
			----*Percent* ----	
MN 'Cash Rent Paid'	39	48	p=0.3139	
MN 'Other income'	51	85	p=0.0000	*
MN 'Rent Received'	84	77	p=0.0774	
MN 'EMV of production inputs – beginning of the year	44	48	p=0.0266	*
MN 'EMV of production inputs – end of the year	47	47	p=0.1107	
MN 'EMV of trucks and cars'	14	20	p=0.2198	
TX 'Cash Rent Paid'	60	61	p=1.0000	
TX 'Other income'	75	95	p=0.0000	*
TX 'Rent Received'	92	91	p=1.0000	
TX 'EMV of production inputs – beginning of the year	83	75	p=0.2252	
TX 'EMV of production inputs – end of the year	81	74	p=0.3255	
TX 'EMV of trucks and cars'	31	15	p=0.0041	*

* significant that the $\alpha < 0.05$ level

Table 4: Percentage of item nonresponse for
2003

State/Variable	Short Form %	Long Form %	p value	
	----*Percent* ----			
MN 'Cash Rent Paid'	42	43	p=0.9013	
MN 'Other income'	57	80	p=0.0002	*
MN 'Rent Received'	87	91	p=0.3550	
MN 'EMV of production inputs – beginning of the year	48	27	p=0.0000	*
MN 'EMV of production inputs – end of the year	47	23	p=0.0000	*
MN 'EMV of trucks and cars'	20	13	p=0.3174	
TX 'Cash Rent Paid'	54	55	p=0.8011	
TX 'Other income'	93	94	p=0.5086	
TX 'Rent Received'	93	96	p=0.1629	
TX 'EMV of production inputs – beginning of the year	71	70	p=0.1096	
TX 'EMV of production inputs – end of the year	70	73	p=0.0801	
TX 'EMV of trucks and cars'	19	15	p=0.3623	

* significant that the $\alpha < 0.05$ level

Examples of changes made during the redesign

<u>1. Demographics</u>

Changes made:
- Background color used to highlight answer boxes.
- Check boxes instead of coding used for multiple choice questions.
- Key codes moved outside the answer boxes and put in a lighter font.

2001 Short Form

2003 Short Form

Appendix A

2. Farm Debt

Changes made:
- Skip instructions more clearly identified with arrows and text.
- Yes/No answers directly lined up vertically.
- Use of color instead of lines to identify cells in the table.

2001 Short Form

2003 Short Form