Effect of Internet Response Mode Designs on Data Quality and Ease of Use

Kelly D. Allmang, Kevin J. Zajac
U.S. Census Bureau, Washington DC 20233

Abstract [1]

The Census Bureau recently conducted the 2005 National Census Test, a multi-faceted test that explored questionnaire content and design alternatives. One objective was to test two designs of the 2005 Census Internet data collection instrument: a person-based approach and a topic-based approach. The person-based approach resembled that of a traditional paper census questionnaire, where the respondent provided data for one person at a time. In the topic-based design, the respondent provided person-level data for the entire household one question at a time. In this paper, we compare the two Internet design approaches. The overall goal of this testing was to determine how best to enhance the Internet application to take advantage of the technology while maintaining or improving the data quality.

Keywords: Topic-based, Person-based, Census

1. Background

In preparation for the 2010 Census, the Census Bureau conducted the 2005 National Census Test (NCT); Census Day was September 15, 2005. The test objectives were multifold: to improve completeness and accuracy of responses, reduce respondent and data capture errors for all response modes, improve accuracy of population coverage, determine operational feasibility of replacement questionnaires, and improve self-response using a bilingual questionnaire. In an effort to reduce respondent and data capture errors, as well as to improve respondent friendliness of the questionnaires, one aspect of the test was focused on the design of the Internet response mode. Since the Internet already proved to be a doable response mode for collecting census data in the 2003 National Census Test (Brady, Stapleton and Bouffard, 2004), the overall goal of this study was to determine how best to enhance the Internet data collection application to take further advantage of the technology while maintaining, or improving, the data quality.

---

[1] This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

In an attempt to enhance the Internet application, the overall presentation of the application was explored. In the 2003 NCT, the Internet application resembled the paper questionnaire. However, another automated census instrument, used for nonresponse follow-up, utilized a grid-like, topic-based design. Thus, one objective of this test was to compare two designs of the 2005 Census Internet application: a person-based approach and a topic-based approach. The person-based approach resembled that of a traditional paper questionnaire, where the respondent provided data for one person at a time. In the topic-based approach, the respondent provided person level data for each person one item at a time. With either design, the Internet application differed from the paper questionnaire. There were multiple versions of the paper questionnaire that tested experimental treatments. Since a single set of questions was used to analyze the Internet application, these questions did not match a specific version of the paper questionnaire. Likewise, with the functionality of the Internet, the navigation through the Internet application differed from the paper questionnaire. This paper compares the two Internet design approaches. Self-response rates, item nonresponse rates, and satisficing (respondent offers responses that seem reasonable without any memory search or integration) serve to potentially indicate data quality issue, while completion time and break-off rates compare the ease of use for the two Internet application designs.

2. Methodology

2.1 Panel Design

The 2005 NCT consisted of 20 experimental panels, with a total of 420,000 sampled housing units. The experimental treatments explored variations of questionnaire content (including race/ethnicity, tenure, relationship, age/date of birth), questionnaire design alternatives, strategies for implementing a replacement questionnaire mailing, methods to improve within household coverage, and a bilingual Spanish census form.

2.2 Mailing Strategy

The mailing strategy for the 2005 NCT consisted of multiple mailings to contact the sampled housing units. Every housing unit was sent an advance letter. The advance letter stated that households had been selected

to participate in the 2005 NCT and would soon receive a request to complete a questionnaire for the 2005 National Census Test.

The second mailing was the questionnaire package. Each housing unit received a paper questionnaire, along with a first-class postage-paid return envelope. A letter from the Census Bureau's Director was also included within the mailing package. The letter not only encouraged the households to respond, but it also provided households with the opportunity to respond via the Internet. The Uniform Resource Locator (URL) and the 18-digit housing unit identification number (ID), which were necessary to access the website, were also provided within this mailing package.

The third mailing consisted of a reminder postcard. The reminder postcard included a statement reminding the households respond to the census test, if they had not already done so, and thanking those who had already responded. The reminder postcard also provided instructions so that the household could respond via the Internet.

The fourth, and final, mailing was the targeted replacement mailing. The replacement mailing was sent to all housing units that did not respond prior to September 13, 2005. In all but one panel, the replacement mailing package resembled that of the initial questionnaire package (i.e. a questionnaire identical to the initial questionnaire, first-class postage-paid return envelope, and a letter from the Census Bureau's Director urging response and presenting the opportunity to respond via the Internet). In the remaining panel, the replacement mailing consisted of a letter encouraging the household to respond via the Internet or to complete and return the initial questionnaire.

2.3 Response Modes

All housing units selected for the 2005 NCT were presented the opportunity to respond via two modes. The two modes were paper and Internet.

2.3.1 Paper

The 2005 NCT questionnaire allowed the respondent to list names for up to 12 household members. For up to six household members, the questionnaire provided space for reporting person-level data items (name, relationship, sex, age/date of birth (DOB), Hispanic origin, race, ancestry, and a probe to assess the accuracy of coverage). A subset of these items was collected for an additional six members. Also, the questionnaire collected housing unit data items (household population count, tenure, and a probe to assess the accuracy of coverage).

2.3.2 Internet

The 2005 Internet application collected the same data items as the paper questionnaire. For the 2005 NCT, the wording, presentation, and functionality of the Internet instrument needed to diverge from paper to better suit the electronic response mode. Moreover, multiple versions of questions were being tested in the 2005 NCT. However, due to cost and implementation constraints, only one version could be chosen for the Internet. Thus, specific question versions were selected based on input from subject matter experts, who had a sense of the question version that might move forward for the 2010 Census. Hence, the content of the Internet application did not exactly match any of the paper questionnaires. Households that chose to respond via the Internet were no longer considered part of their original experimental panel, since they were not exposed to the experimental treatment. Thus, they were removed from the subsequent treatment analysis and only included in the Internet analysis.

The Internet application consisted of two different presentations of the questions: a person-based approach and a topic-based approach. The person-based approach resembled the paper questionnaire and collected data for each person, one person at a time. The topic-based approach collected data for the entire household, one item at a time. Each housing unit in the sample was pre-assigned to either the topic-based or person-based design. [The housing units were sorted by their housing unit identification. The first housing unit was randomly assigned to one of the designs. The designs were then alternately assigned to every other housing unit.] When the respondent entered their ID to access the Internet application, the system checked the topic or person flag to serve the appropriate path.

Both designs of the Internet application incorporated several interactive features. Questions were tailored to specific persons by incorporating the name, if given, into the question. That is, instead of "What is this person's sex?" the question read, "What is Mary's sex?" Secondly, tabs were added at the top of the screen for navigational purposes and as progress indicators. However, respondents could not select a tab that they had not previously accessed, as the respondent moved through the application in a linear fashion. Thus, the respondent was forced to see every question before advancing. The application incorporated soft edits, which alerted respondents to incomplete or invalid responses, but did not require

corrections to these items. In addition to tabs and soft edits, help links and a "Privacy Policy" link were included on every page. Once respondents entered data for all household members, they could view a summary of their answers prior to making a final submission.

2.4 Sample Design

The 420,000 housing units, selected from mailout/mailback areas of the country, were proportionately allocated into two strata that reflect anticipated differences in the race and ethnic composition and are believed to be associated with response propensity of the mailout/mailback universe. The high non-White or Hispanic concentration stratum, which encompassed roughly 32% of the total mailout/mailback universe, contained a high proportion of the non-White and Hispanic populations. The remaining 68% of the housing units fell in the low non-White or Hispanic concentration stratum. Estimates presented in this paper were weighted to account for oversampling of the high non-White or Hispanic concentration stratum.

2.5 Calculation of Self-Response Rates

The self-response rate is a measure of respondent behavior with regard to responding to a census test. The denominator is the number of sampled housing units after removing those cases identified by the United States Postal Service (USPS) as "undeliverable as addressed" (UAA). A case was defined as UAA when there was no response for that housing unit and the USPS flagged both the initial questionnaire and the replacement questionnaire mailings as UAA (Rothhaas, 2005b). Any housing unit flagged as UAA was considered an ineligible unit.

The numerator of the self-response rate is the number of sampled housing units, for which a nonblank, primary return was received. A return was defined as blank if fewer than two census items were "completed". These items verified for completeness include name, relationship, sex, age or date of birth, Hispanic origin, race, and ancestry (Rothhaas, 2005a). When more than one return was received from a single housing unit, based on the housing unit ID, the first nonblank return received (based on date and time received) was selected as the primary return for data analysis. In the rare case that two nonblank paper forms were checked in on the same day and in the same batch, then the initial questionnaire was selected as the primary. If both a nonblank paper return and nonblank Internet return were received for the same housing unit on the same day, the nonblank paper

return was selected as the primary return for data analysis (Rothhaas, 2005c). The formula for the self-response rate is given below.

$$Self\text{-}response\ rate = \frac{\#\ of\ nonblank,\ primary\ returns}{panel\ sample\ size - UAA\ for\ the\ panel} * 100$$

Please note that this self-response rate corresponds to the rates used in previous census tests, including the 2003 NCT, the Census 2000 experiments, and the 1992 and 1993 Census Tests. We use this self-response rate because it is not subject to variation in UAA rates. Specifically, the denominator of the self-response rate excludes cases for which eligibility cannot be determined, such as units that are UAA. Therefore, any variation in the UAA rates across panels will not contribute to differences in the self-response rates.

Lastly, please note that the self-response rate defined here is not comparable to the Census 2000 mail response or mail return rates. The self-response rate is not a return rate in the sense that we do not definitively know the occupancy status of housing units included in the denominator or the status of cases that are excluded as UAAs.

2.6 Calculation of Item Nonresponse Rates

Item nonresponse rates are an indicator of data quality. They refer to the percentage of records with missing data for a particular item. This analysis is restricted to Internet returns for a housing unit. The item nonresponse rates are calculated as follows:

$$Item\ nonresponse\ rate = \frac{\#\ of\ records\ with\ missing\ data\ for\ a\ particular\ item}{total\ number\ of\ nonblank\ primary\ Internet\ return\ records} * 100$$

Item nonresponse rates were calculated for both housing unit level items and person level items. For the housing unit level items, the term "records" referred to housing units. The total number of housing units was defined as the number of housing units from all Internet returns. For person-level items, the term "records" referred to persons. The total number of persons was defined as the number of data defined persons listed on all Internet returns. A data defined person record had at least two entries that met specified completion criteria for the following items: name, relationship, sex, age/date of birth, Hispanic origin, race, and ancestry (Reiser, 2005).

For this analysis, item nonresponse rates were calculated for seven person level items (relationship, sex, age/date of birth, Hispanic origin, race, ancestry, and coverage overcount) and three housing unit level

items (population count, tenure, and coverage undercount).

2.7 Calculation of Break-Off Rates

Break-off rates are the percentage of primary returns that started the questionnaire, but did not submit the questionnaire. These break-off rates may, or may not, be indicators of the application's ease of use. The numerator is the difference between the total number of Internet returns and the number of successfully submitted Internet returns. A return is considered successfully submitted if the respondent clicked submit on the review/submit screen at the end of the instrument. The denominator of the break-off rate is the total number of Internet returns.

2.8 Satisficing

Satisficing occurs when the respondent offers responses that seem reasonable without any memory search or integration. For several person-level data items, the topic-based design displayed the questions as matrices. Thus, there was a fear of the respondent simply selecting all checkboxes down one column without reading all of the answer categories and selecting the most appropriate. The race variable was studied to support/refute the satisficing claim, as it offered many response options and was not as straightforward as other questions.

2.9 Variance Estimation

We used Wesvar to compute standard errors for all estimates using a stratified jackknife replication procedure to account for the stratification in the sample. The housing units were sorted in the same order that they were selected and clusters of housing units, or housing units selected at each hit, were sequentially assigned to one of the 250 replicates. This assignment approach accounted for the clustering of persons with a household in computing errors for person level estimates, since persons within households are clustered together in the same replicate.

3. Results

3.1 Self-Response Rate Analysis

Table 1 illustrates that, nationally, the person-based and topic-based Internet application self-response rates are comparable. Excluded from the response rate calculations are any returns from the Internet encouragement panel, where the replacement package included a letter encouraging response by Internet in lieu of sending a replacement questionnaire.

Table 1. Self-Response Rates at the National Level.

	Estimate	S.E.
Total	60.8%	0.14%
Internet Application	7.2%	0.06%
Person-Based Design	3.6%	0.06%
Topic-Based Design	3.6%	0.06%
Paper	53.6%	0.14%

3.2 Item Nonreponse Analysis

Table 2. Household Level Item Nonresponse Rates

	Person-based design		Topic-based design	
	Estimate	S.E.	Estimate	S.E.
Population Count*	-	-	-	-
Tenure**	0.07%	0.05%	0.20%	0.05%
Coverage Overcount	0.05%	0.03%	0.04%	0.02%

*Respondent had to answer the question in order to continue
** Significant difference between the two designs at alpha=0.10

Table 2 displays the household level item nonresponse rates. There were no estimates for the population count nonresponse. In order to continue with the Internet application, the respondent had to answer the household population count question.

The only household level data item that had a statistically significant difference in nonresponse rates between the person-based design and the topic-based design was tenure. The person-based Internet application design had a statistically lower item nonresponse rate. Figures 1 and 2 show screen shots of the person-based and topic-based household screens respectively. On the person-based household screen, the tenure question was the third question on the screen. In addition, the tenure question immediately preceded the "next" button, which the respondent would have clicked in order to continue on with the application. For the topic-based design household screen, the tenure question was still the third question on the screen. However, the question did not precede the "next" button. Instead, it was buried in the middle of the screen. Thus, the screen density and question placement could have caused the topic-based tenure nonresponse rate to be higher than the person-based.

At first, for the person-level item nonresponse rates, the person-based design rates appeared to be significantly lower than the topic-based rates, see

Table 3. In addition, there was a pattern to the topic-based item nonresponse rates. The data items further into the application had progressively higher item nonresponse rates. This implied that the item nonresponse rates were compounded by break-offs and the definition of data-defined persons. Thus, item nonresponse rates could be analyzed by components: break-offs and then item nonresponse for successfully submitted returns [returns for which the respondent clicked the submit button].

Table 3. Item Nonresponse Rates for Internet Returns

Item	Person-Based Application		Topic-Based Application	
	Estimate	S.E.	Estimate	S.E.
Relationship*	0.03%	0.02%	0.01%	0.01%
Sex	0.21%	0.03%	0.73%	0.10%
Age/ Date of Birth	0.60%	0.06%	1.12%	0.11%
Hispanic Origin	0.60%	0.07%	1.55%	0.15%
Race	0.92%	0.09%	1.92%	0.17%
Ancestry	0.99%	0.08%	2.51%	0.21%
Coverage Overcount	0.78%	0.06%	2.49%	0.20%

* Relationship excludes Person 1.

3.2.1 Break-off Rates

The person-based application had significantly more break-offs than the topic-based application. [The person-based design had a 5.04% (0.22% S.E.) break-off rate, while the topic-based design had a 3.74% (0.20% S.E.) break-off rate.] With this significant difference, it is important to study where the break-offs occur. The location of the break-off could have impacted whether or not a person was deemed data-defined or not. Recall that a data-defined person was a person record with at least two "completed" data items.

Roughly 57 percent of the person-based design break-offs occurred at or before the person one residence screen. Thus, the respondent would have already seen the name, relationship, sex, date of birth, age, Hispanic origin, race, and ancestry screens. Thus, person one could potentially be data-defined. However, no information was collected for on any additional persons within the household.

Approximately 88 percent of the topic-based design break-offs occurred at or before the residence (overcount) screen. At this point, the respondent would have seen the name, relationship, sex, date of birth, age, and Hispanic origin data item for all persons with in the household. Thus, all persons within the household could potentially be data-defined.

Table 4 displays the effects of the break-offs. From the first question on the Internet application, the household population count question, there was no significant difference in the reported mean household size. However, when the mean number of data-defined persons per household was calculated, the topic-based Internet application resulted in a significantly larger mean. This significant difference carried over into the mean household size discrepancy (difference between the reported and calculated). Most likely, the difference in the mean number of data-defined persons between the person-based and topic-based Internet designs was due to break-offs.

Table 4. Household Population Counts by Internet Application, National Level

Factor	Person-Based Application		Topic-Based Application	
	Estimate	S.E.	Estimate	S.E.
Reported Mean Household Size	2.69	0.02	2.68	0.02
Calculated Mean Household Size*	2.60	0.02	2.65	0.02
Mean Household Size Discrepancy*	0.08	0.01	0.03	0.00

*Significant difference between the two Internet applications at alpha=0.10

In order to account for this difference, only successfully submitted Internet records will be considered in the remaining analysis. Recall, a return was considered successfully submitted if the respondent clicked submit on the review/submit screen.

3.2.2 Successfully Submitted Returns

Table 5 shows item nonresponse rates for person level items by application for all persons. Generally, there was no significant difference in item nonresponse rates by Internet application. Only the coverage overcount item was significantly different between the two applications. In the coverage overcount, the wording could account for the difference in the item nonresponse rates. [The screen shots of the coverage overcount screens are displayed at the end of the paper,

Figures 3 and 4.] For the person-based approach, the overcount question asks, "Does Person 1 sometimes live or stay somewhere else?" In order to adjust to the topic-based approach, the question now reads, "Do any of these people live or stay somewhere else?" The question wording for the topic-based design implied a single yes/no answer and the wording " Do any of these people…" differed from the other question wording within that design. Thus, it was thought that the item nonresponse rate for the topic-based design was high due to the respondents answering the question only for person one. However, analysis was conducted and the hypothesis was not realized.

Table 5. Person Level Item Nonresponse Rates for Successfully Submitted Internet Returns

Item	Person-Based Application		Topic-Based Application	
	Estimate	S.E.	Estimate	**S.E.**
Relationship *	0.02%	0.02%	0.01%	0.01%
Sex	0.06%	0.03%	0.14%	0.05%
Age/ Date of Birth	0.35%	0.06%	0.38%	0.07%
Hispanic Origin	0.24%	0.06%	0.21%	0.05%
Race	0.44%	0.08%	0.37%	0.07%
Ancestry	0.31%	0.06%	0.36%	0.07%
Coverage Overcount+	0.06%	0.02%	0.12%	0.03%

Relationship excludes Person 1.
+Significant difference between the two Internet applications at alpha=0.10

3.3 Average Completion Time

Average completion times are only computed for successfully submitted Internet returns, those returns where the respondent clicked submit on the review/submit screen. There was no significant difference between the two designs in the average length of time it took to complete the Internet application. The person-based design took 7.08 minutes where the topic-based design took about 6.87 minutes.

3.4 Satisficing Analysis

The two Internet application designs differed not only in question wording but also visual presentation. Several questions concerning person-level data items

were displayed as matrices. The race question was one such question, where the rows represented the persons within the household and the columns represented the response options. Thus, for the topic-based design, there was a fear that a respondent would simply click down one column.

In order to analyze this, we studied an indicator of whether there were any differences in race among household members. The person-based design yielded 7.99% (0.27% S.E.) of households with a diff.erence in race, whereas the topic-based design yielded 8.01% (0.27% S.E.) of households with a difference in race among household members. Therefore, there was no statistical difference between the two designs in the percentage of households whose members reported a difference in race. Thus, for the race variable, the fear of satisficing was not realized.

4. Conclusions

There is no significant difference between the two designs in terms of respondent burden, or the length of time it takes to complete the application. Indicators of data quality, such as item nonresponse rates and satisficing, are also comparable across the two designs. However, the person-based design has a statistically lower tenure nonresponse rate than the topic-based design. This difference is hypothesized to be caused by the screen density and question placement. In addition, the probe assessing the coverage overcount is also statistically different between the two designs. Once again, the topic-based design had a statistically higher overcount nonresponse rate.

The comparison of the two Internet application designs yields a trade-off. The topic-based design produced more data-defined persons, however not necessarily with complete data. On the other hand, the person-based design produced fewer data-defined persons. However, these persons have more complete data. Further research and discussion is needed to decide the relative importance of these issues.

References

Brady, S., Stapleton, C., and Bouffard, J. (2004) "2003 National Census Test: Response Mode Analysis", Internal Census Bureau Memorandum, September 29, 2004.

Reiser, C. (2005) "Definition of Data Defined Persons for the 2005 National Census Test", Internal Census Bureau memorandum, May 31, 2005.

Rothhaas, C. (2005a) "Definition of Blank Forms for the 2005 National Census Test", Internal Census Bureau Memorandum, May 31, 2005.

Rothhaas, C. (2005b) "Definition of Undeliverable As Addressed Flag for the 2005 National Census Test", Internal Census Bureau Memorandum, March 15, 2005.

Rothhaas, C. (2005c) "Duplication Return Rule for the 2005 National Census Test", Internal Census Bureau Memorandum, May 10, 2005

Figure 1. Person-Based Household Screen

Figure 2. Topic-Based Household Screen

Figure 3. Person-Based Application Coverage Overcount Screen

Figure 4. Topic-Based Application Overcount Screen