A Comparison of Closed- and Open-Ended Question Formats for Select Housing Characteristics in the 2006 American Community Survey Content Test[*]

John Chesnut, Decennial Statistical Studies Division, U.S. Bureau of the Census
Jeanne Woodward, U.S. Census Bureau; Ellen Wilson, U.S. Census Bureau
Washington, D.C. 20233

Keywords: closed- and open-ended response, item nonresponse, reliability, systematic response error

1. Introduction

In January through March 2006, the U.S. Census Bureau conducted the first test of new and modified content for the American Community Survey (ACS) since the survey reached full implementation levels of data collection. The results of that testing will help determine the content for the 2008 American Community Survey[1]. One of the research objectives of this test was to conduct an experimental study of the impact of using open-ended question formats compared to closed-ended question formats for three different housing questions – property value, number of vehicles kept by members of the household, and number of rooms and bedrooms in the household. The three questions require varying amounts of knowledge to respond and thus may differ in how respondents use an open- versus closed-ended response format. For each of the three items, this paper examines the differences in data quality resulting from the two different response formats in terms of item nonresponse, reliability, and systematic response error.

The survey methodology literature points out the advantages and disadvantages to using closed-ended versus the open-ended question format largely in the context of attitudinal or public opinion questions (Schuman and Presser, 1981; Sudman, Bradburn, and Schwarz, 1996; Converse and Presser, 1986). Disadvantages to the closed-ended format include biasing the respondents by the given response options. For example, Bradburn and Sudman (1979) found that the presence of low frequency options for drinking and sexual activity influenced the respondents to report lower frequencies for these activities. In addition, the closed-ended format may make the respondent feel limited with the available response choices and choose not to report an answer. The closed-ended format does have its advantages. For example, the closed-ended format response categories can give clues to the respondent on how to interpret the researcher's intended meaning of a question. In addition, open-ended formats may be at a disadvantage in the case where respondents do not provide enough detail to meet the researcher's objectives. Furthermore, closed-ended responses may require less coding and processing after collecting the data.

Traditionally, the ACS has used closed-ended formats for the housing questions included in this study. The motivation for testing whether the ACS can use open-ended formats for these questions varied by question. All of the questions when converted to the open-ended format will reduce the amount of questionnaire space required for each question. In the case of the property value question, the open-ended version will in theory allow the respondent to provide a more "precise" response. Economist and housing analysts at the Housing and Urban Development (HUD) have expressed that they have encountered difficulty using the bracketed data, and have recommended that the ACS collect property value information as a write-in rather than continuing with the current categorical approach established in prior decennial census data collection efforts. Furthermore, the categories used in Census 2000 may not serve them well in the coming years if the housing market continues at the pace established in the first half of this decade. In addition, this is the only dollar value on the ACS questionnaire that is currently collected as categorical data. The categorical property value data is difficult to inflation-adjust from year to year.

Historically, the property value question has been asked as both a closed and open-ended question in the decennial census. The closed-ended format has been used since the 1960 census; prior to that it was asked as a write-in. In the 1960 census, a 10 category response option was used ranging from "Less than $5000" to "$35,000 or more." In the 1970 census there were 11 categories ranging from "Less than $5,000" to "$50,000 or more." The 1980 question increased the categories to 24 and in 1990 the categories were increased to 26. In 2000 we dropped back down to 20 categories. For each census the property value ranges were adjusted to reflect the continuing appreciation in housing prices.

Reviewing the history of the vehicles question, we find that the vehicles question has been included in the census since 1960 and has always been asked as a closed-ended question. A four category response option was used ranging from "none" to "three or more" cars.

Finally, we review the history of the rooms and bedrooms questions. In the 1940 and 1950 census, the rooms question was asked as an open-ended question. The bedrooms question was first asked in 1960, and since 1960 both the rooms and bedrooms questions have been asked as closed-ended questions.

Cognitive pre-testing for the ACS content test gave some insight into which response format respondents may prefer. The cognitive pre-testing showed that respondents preferred the closed-ended version to the open version for the property value question. For the property value, participants expressed that giving an exact property value in the open version was more difficult and burdensome than giving a range for their property value in the closed version. As for the vehicles and rooms/bedrooms questions, the respondents gave no suggestion as to whether the open- or closed-ended version of the question was better (Kerwin et al., 2005).

In addition to changing the question format from a closed- to an open-ended format, subject matter experts and inter-agency committees developed other changes to improve the question stem, instructions, or examples. These changes were incorporated along with the open-ended format into the "test" question version. The "control" version of the question used the closed-ended layout (see appendix for a facsimile of the control and test questions). Therefore, differences that we observe between the control and test treatment groups may not entirely be attributed to a difference in layouts.

Changes made to the rooms question included the following:
- add the word "separate" to the question stem,
- add an instruction that defines a "room,"
- add an instruction to include bedrooms and kitchens in the count of rooms and an instruction to exclude unfinished basements and drop "half rooms."

Changes made to the bedrooms question included the following:
- add language that explicitly links the total count of rooms and the count of bedrooms,
- provide the heuristic/rule to use for defining a bedroom as part of the instruction, and
- provide an instruction for writing "0" separate bedrooms for efficiency/studio apartments.

Changes made to the vehicles question included the following:
- add the term "SUVs," and
- add an instruction to exclude motorcycles and other recreational vehicles.

Changes made to the property value question included the following:
- change the question stem by dropping the first part of the question referring to value of the property and specifying who should include their lot value in their response.

2. Methods

2.1.1 The 2006 ACS Content Test data collection

The 2006 ACS Content Test consisted of a national sample of approximately 62,900 residential addresses in the contiguous United States (the sample universe did not include Puerto Rico, Alaska and Hawaii). To meet the primary test objective of evaluating changes to the question wording, approximately half of the sample addresses were assigned to a test group and the other half to a control group. For the topics already covered in the ACS, the test group included the proposed alternative version of the questions, and the control group included the current version of the questions as asked on the ACS. For the property value, rooms, bedrooms, and vehicles questions, the control version was designated as the closed-ended format, and the test version was designated as the open-ended format.

The ACS Content Test used a similar data collection methodology as the current ACS, though cost and time constraints resulted in some deviations. Initially, the ACS collects data by mail from sampled households, following a mailing strategy geared at maximizing mail response (i.e., a pre-notice letter, an initial questionnaire packet, a reminder postcard, and a replacement questionnaire packet). The Content Test implemented the same methodology, mailing each piece on the same dates as the corresponding sample panel in the ACS. However, the Content Test did not provide a toll-free number on the printed questionnaires for respondents to call if they had questions, as the ACS does. The decision to exclude this service in the Content Test primarily reflects resource issues in developing the materials needed to train and implement the operation for a one-time test. However, a benefit of excluding this telephone assistance is that it allows us to collect data that reflects the respondent's interpretation and response without the aid of a trained Census Bureau interviewer.

The ACS follows-up with mail nonrespondents first by Computer Assisted Telephone Interviewing (CATI) if a phone number is available, or by Computer Assisted Personal-visit Interviewing (CAPI) if the unit cannot be reached by mail or phone. For cost purposes, the ACS subsamples the mail and telephone nonrespondents for CAPI interviewing. In comparison, the Content Test went directly to CAPI data collection for mail nonrespondents, dropping the CATI data collection phase in an effort to address competing time and resource constraints for the field data collection staff. While skipping the CATI phase changes the data collection methods as compared to the ACS, eliminating CATI allowed us to meet the field data collection constraints while also maintaining the entire mail nonrespondent universe for possible CAPI follow-up. Using CATI alone for follow-up would have excluded households for whom we did not have a phone number.

The ACS also implements an edit procedure on returned mail questionnaires, identifying units for follow-up who provided incomplete information on the form, or who reported more than five people living at the address (the ACS questionnaire only has space to collect data for five people.) This is called the Failed Edit Follow Up operation (FEFU). The ACS calls all households identified as part of the FEFU operation to collect the remaining information via a CATI operation. The Content Test excluded this follow-up operation in favor of a content reinterview, called the Content Follow-Up (CFU). The CFU also contacts households via CATI, but the CFU serves as a method to measure response error, providing critical evaluative information. The CFU operation included all households who responded by mail or CAPI and for whom we had a phone number. More information about the CFU operation follows below.

The Content Test mailed questionnaires to sampled households around December 28, 2005, coinciding with the mailing for the ACS January 2006 sample panel. The Content Test used an English-only mail form but the automated instruments (both CAPI and CFU) included both English and Spanish translations. Beginning in February 2006, a sample of households that did not respond by mail was visited by Census Bureau field representatives in an attempt to collect the data. The CAPI operations ended March 2, 2006.

2.1.2 Content Follow-Up data collection

The CFU reinterview, conducted by the Census Bureau's three telephone centers, provided a method for measuring response error. About two weeks after receiving returned questionnaires or completed CAPI interviews, all responding units entered the CFU operation. Telephone staff completed the CFU interviews between January 17 and March 17, 2006. At the first contact with a household, interviewers asked to speak with the original respondent.

If that person was not available, interviewers scheduled a callback at a time when the household member was expected to be home. If, at the second contact, we could not reach the original respondent, interviewers completed the interview with another adult household member.

The CFU reinterview did not replicate the full ACS interview. Rather, the CFU used the roster and basic demographic information from the original interview and only asked questions specific to the analytical needs of the Content Test. Reinterview questions were of two general formats: the same question as asked in the original interview (in some cases, modified slightly for a CATI interview), or a different set of questions providing more detail than the question(s) asked in the original interview for the same topic. For topics in which the CFU asked the same question as the original interview, the CFU asked the test or control version of the question based on the original treatment. For these cases, the goal was to measure the reliability of the answers – how often we obtained the same answer in the CFU as we did in the original mail or CAPI data collection. For topics using a different question or set of questions than the original interview, we asked the same detailed series of questions regardless of the original treatment condition. Generally, these questions were more numerous than what we could ask in the ACS. For the topics covered in this report, the goal was to measure how close the original answers were to the more detailed CFU answers.

Content Follow-up for the property value question was intended to be a simple re-ask of the original question. However, the control version was not a "true" re-ask for the CAPI respondents since the CAPI instrument was an open-ended question with no instructions for the interviewer to reference the property value ranges defined in the control mail version. This was a result of our decision to use the current production CAPI instrument for the control version. In production, the CAPI instrument is designed such that the response format for the property value question is open-ended, which differs from the close-ended format used for the production mail questionnaire. Therefore to work around this limitation, we restricted our analysis of the CFU property value data to those respondents who responded to the content test via the mail questionnaire.

Table 1. Property Value Response Formats by Mode

Mode	Control Panel	Test Panel
MAIL	Closed	Open
CAPI	Open	Open
CFU	Open with Instruction*	Open

* Interviewer instructed to read categories, if needed

The CFU approach for the rooms and bedrooms questions was different than a straight "re-ask" of the rooms and bedrooms questions. Our objective was to gain a "better" measure of the rooms and bedrooms count. We asked a series of questions about the functional use of specific rooms similar to the method followed in the American Housing Survey. This approach allowed us to filter out bathrooms and other areas within housing units that should not be included in the count of rooms. Note that the vehicles question was not included in the CFU study.

2.2 Sample Design

The sample design for the ACS Content Test consisted of a multi-stage design, with the first stage following the Census 2000 Supplementary Survey (C2SS) design for the selection of Primary Selection Units (PSUs) defined as counties or groups of counties. The first stage selection of PSUs resulted in 413 PSUs or approximately 900 counties being selected.

Within sampled PSUs, households were stratified into high and low response area strata based on tract-level mail response rates to the Census 2000 long form, and a stratified systematic sample of households was selected. The strata were defined such that the high response stratum contained 75 percent of the housing units that reside in tracts with the highest mail response rate. The balance of the tracts was assigned to the low response stratum. To achieve similar expected number of mail returns for the high and low response strata, 55 percent of the sample was allocated to the low response strata and 45 percent to the high response strata.

A two-stage sampling technique was used to help contain field costs for CAPI data collection. The initial sample of PSUs was sorted by percentage of foreign-born population, since the majority of that target population will end up responding via CAPI. At least one item undergoing testing in the content test required an adequate sample of this

population. The 20 PSUs with the highest percentage of foreign-born population were included with certainty and the remaining PSUs were sampled at a rate of 1 in 3. For the second stage, mail nonresponding households were sampled at a rate of 1 in 2 within the top 20 PSUs and at a sampling rate of 2 in 3 within the remaining PSUs. The final design designated 151 PSUs for inclusion in the CAPI workload.

In the majority of PSUs, we assigned cases to both the control and test groups. To maintain field data collection costs and efficiencies, PSUs with an expected CAPI workload of fewer than 10 sampled addresses had all of their work assigned to only one treatment (either control or test). The PSUs were allocated to the two groups such that the aggregated PSU characteristics between the two groups were similar for employment, foreign born, high school graduates, disabled, poverty status, tenure, and Hispanic origin. For more information on the 2006 ACS Content Test sample design, see Asiala and Navarro (2006).

There was no sampling for CFU. A CFU interview was attempted for all households responding to the Content Test for which we had a phone number.

2.3 Statistical Methods

To study the impact of using the open-ended question formats versus the closed-ended question formats for the three different housing questions, we conducted statistical tests to determine which of our defined statistical measures were significantly different between the control and test treatment groups. In the case of testing whether a given response distribution was dependent on the question version, we used an adjusted Pearson chi-square test statistic to account for the complex sample design, testing at the 10.0 percent significance level. The Pearson chi-square test statistic was adjusted using the Rao-Scott first order correction (Rao and Scott 1981, 1984). For the remaining analysis, we calculated the difference between the control and test sample estimates then used a two-sided t-test at the 10.0 percent significance level to determine those differences that were significant. Note that all statistical tests performed in this paper use a 10.0 percent significance level to meet Census Bureau policy. All analysis for this paper was performed using WesVar statistical software. Variances used in our statistical tests were estimated with WesVar using the Jackknife variance estimation method.

3. Results

3.1 Response to the Content Test and Content Follow-up

Control and test treatments groups obtained equivalent response rates overall, and for each mode of collection. The table below gives the weighted response rates for each data collection operation and a test of differences between the control and test groups. The overall response rate reflects the final response to the initial data collection (mail and CAPI only). There were no significant differences between response rates for the control and test groups. Note that the denominator for each calculation included only eligible cases for each mode.

Table 2. Content Test Response Rates, Control vs. Test

Response Rate	Control (%)	Test (%)	Difference (%)	Margin of Error (%)	Significant
Overall response rate	95.8	95.5	-0.3	± 0.9	No
Mail response rate	51.5	51.2	-0.3	± 2.2	No
CAPI response rate	92.6	92.1	-0.4	± 1.7	No
CFU response rate	75.9	76.4	0.5	± 1.6	No

3.2 Rooms and Bedrooms

A research objective common across all of the topics that we tested in the content test was to determine whether the changes being tested improved or maintained the levels of item missing data produced by the control question version. To determine the effect of the test question version on missing data, we compared the item nonresponse rates (INR), the proportion of household or person responses with "missing data," between the control and test treatment groups. Note that the definition of missing data varied by question. For the rooms and bedrooms control

versions, a nonresponse was defined as no check box checked for the rooms and bedrooms count categories. For the test version, a nonresponse was defined as no entry or an illegible entry in the write-in field.

Table 3 shows no significant differences in the item nonresponse rates between the control and test versions for both the rooms and bedrooms questions at the national level and for the high response areas (HRAs). However, for low response areas (LRAs) we observe that the test version of both the rooms and bedrooms questions resulted in marginally significant increases in the nonresponse rate. Based on these results we conclude that the test version maintains the level of missing data produced by the control version.

Table 3. Item Nonresponse Rates for Rooms/Bedrooms Questions

Strata	Control (%)	Test (%)	Difference (%)	Margin of Error (%)	Significant
Rooms					
National	4.1	4.8	0.7	$\pm$ 0.9	No
High Response Area	3.9	4.6	0.6	$\pm$ 1.2	No
Low Response Area	4.6	5.5	1.0	$\pm$ 0.9	Yes
Bedrooms					
National	3.4	4.3	0.8	$\pm$ 1.0	No
High Response Area	3.3	3.9	0.7	$\pm$ 1.2	No
Low Response Area	4.0	5.3	1.3	$\pm$ 0.9	Yes

Based on results from the Census 2000 Content Reinterview Survey, subject matter experts hypothesized that respondents were under-reporting the number of rooms for their housing unit (Singer and Ennis 2003). To address this problem, the subject matter experts and interagency committee proposed changes to reduce the under-reporting (cf. Section 1 for a listing of the changes). Table 4 shows the median number of rooms reported by responding households as well as the median number of bedrooms for both the control and test versions of the questions. The test panel resulted in a significantly larger median number of rooms. Therefore, the test version reduced the under-reporting of rooms. For the bedrooms question, we observe that the changes to this question did not impact the median number of bedrooms.

Note that the medians for both control and test versions were calculated using a linear interpolation method suitable for use with categorical data. To facilitate this method we associated each room or bedroom category with an interval. For example, a 5-room category now becomes the interval (4.5, 5.5). The median was calculated by first identifying the interval containing the median using a cumulative frequency distribution. Next, we used linear interpolation to determine the placement of the median value between the interval endpoints.

Table 4. Median Rooms and Bedrooms, Control vs. Test

	Control (#)	Test (#)	Difference (#)	Margin of Error (#)	Significant
Rooms	5.3	5.7	0.4	$\pm$ 0.1	Yes
Bedrooms	2.7	2.7	0.0	$\pm$ 0.0	No

Table 5 shows the household room count distribution by control and test. From the chi-square statistic, we find that the rooms distribution is dependent on the question version. Reviewing the individual t-test comparisons, we observe that the test version of the rooms question produces significant increases for one, six, seven, and nine or more room housing units and significant decreases for two, three, four, and five room housing units. More than likely, the "shifting" taking place in the response distribution for the rooms count is not due to changing from a closed- to an open-ended layout, but due to the other changes introduced in the test version of the question.

Table 5. Rooms Distribution, Control vs. Test

Rooms	Control (%)	Test (%)	Difference (%)	Margin of Error (%)	Significant
1	1.5	2.3	0.8	$\pm$ 0.4	Yes
2	3.7	2.1	-1.6	$\pm$ 0.6	Yes
3	9.1	7.7	-1.5	$\pm$ 1.0	Yes
4	17.5	15.7	-1.8	$\pm$ 1.3	Yes
5	22.3	19.0	-3.2	$\pm$ 1.5	Yes
6	17.4	19.1	1.7	$\pm$ 1.2	Yes
7	11.8	13.5	1.7	$\pm$ 1.2	Yes
8	8.2	8.9	0.7	$\pm$ 0.9	No
9 or more	8.5	11.7	3.2	$\pm$ 1.0	Yes
Total	100.0	100.0			

$\chi^2 = 82.6$ with 8 degrees of freedom, significant at the 10.0 percent level

Data included in Table 6 indicate that there were a higher percentage of housing units with "0" bedrooms (efficiency apartments) and a lower percentage of "1-bedroom units" in the test treatment group. Based on this result, we conclude that the efficiency instruction added to the test version produced a shifting of "1-bedroom units" to "0-bedroom units". However, when we reproduce this analysis controlling for the mode of response (mail or CAPI), we find that this effect persists only for the CAPI mode. Therefore, we conclude that the significant increase in efficiencies was a result of an efficiency apartment question included in the test version of the CAPI instrument, not the efficiency instruction added in the mail questionnaire.

Table 6. Bedroom Distribution Rates, Control vs. Test –National

Bedrooms	Control (%)	Test (%)	Difference (%)	Margin of Error (%)	Significant
0	1.3	2.5	1.2	$\pm$ 0.5	Yes
1	11.4	10.2	-1.2	$\pm$ 1.0	Yes
2	28.4	27.4	-1.0	$\pm$ 1.8	No
3	39.9	40.4	0.5	$\pm$ 1.7	No
4	15.0	15.2	0.2	$\pm$ 1.2	No
5 or more	4.0	4.2	0.2	$\pm$ 0.7	No
Total	100.0	100.0			

$\chi^2 = 21.3$ with 5 degrees of freedom, significant at the 10.0 percent level

The net difference rate (NDR) is used when we assume that the Content Follow Up interview, which asks more questions and collects more detailed data about a topic, provides a better measure than the control or test versions of a question. The NDR reflects the net change between the original response and the response given for the more detailed CFU questions. In other words, since we assume the CFU provides better data, the NDR indicates to what extent the test or control version of a question over- or under-estimates the topic (or category) of interest. Relative to the CFU estimate, a NDR with a negative value indicates an under-estimate and a positive value indicates an overestimate. A NDR that does not differ significantly from "0" indicates that the question asked in the original test or control interview produces results similar to the more detailed question set asked in CFU. In other words, the question should not result in a systematic over- or under-estimate of the topic (or category) of interest.

For the purpose of this paper, we compared the NDR calculated for the test group to that of the control group to assess which version of the question resulted in less systematic response error, regardless of whether the error reflected an over- or under-estimate. To show this, we provide the difference of the absolute values of the NDRs. Data included in Table 7 show the difference of the absolute values of the NDRs for the control and test. With the

exception of the "9 or more rooms" category, the test panel collected data that was as accurate or better than that collected in the control panel in terms of systematic response error.

Ad hoc analysis of the NDRs by mode of data collection showed that the improvement in the underreporting of 1-room units (efficiencies) only persists for the cases that went to CAPI. This suggests that the inclusion of the "efficiency" screen in the test version of the CAPI instrument helped reduce the systematic response error for collecting data on 1-room housing units.

Table 7. Rooms - Content Followup Comparison Statistics, Net Difference Rates, Control vs. Test

Rooms	Control vs. CFU (%)	Test vs. CFU (%)	Diff* |T|-|C| (%)	Margin of Error (%)	Significant
1	-1.5	-0.9	-0.7	± 0.6	Yes
2	2.9	1.5	-1.4	± 0.5	Yes
3	2.8	1.4	-1.4	± 1.0	Yes
4	2.4	1.5	-0.9	± 1.4	No
5	2.3	-1.6	-0.7	± 1.8	No
6	-3.7	-2.2	-1.5	± 1.8	No
7	-2.6	-1.7	-0.9	± 1.8	No
8	-1.8	-0.9	-0.9	± 1.3	No
9 or more	-0.8	2.8	2.0	± 1.1	Yes

*Difference of the absolute values of the test and control net difference rates

Table 8 shows that the test version of the bedrooms question either reduces or maintains the level of systematic response error produced by the control version. More specifically, the test version reduces the under-estimation of zero bedroom housing units and the over-estimation of 1-bedroom units.

Table 8. Bedrooms - Content Followup Comparison Statistics, Control vs. Test

	Net Diff Rate				
Bedrooms	Control vs. CFU (%)	Test vs. CFU (%)	Diff* |T|-|C| (%)	Margin of Error (%)	Significant
0	-1.8	-0.7	-1.1	± 0.6	Yes
1	1.7	0.9	-0.8	± 0.6	Yes
2	0.3	0.7	0.4	± 0.8	No
3	-0.4	-1.1	0.7	± 0.9	No
4	0.0	-0.1	0.1	± 0.7	No
5 or more	0.1	0.3	0.1	± 0.4	No

*Difference of the absolute values of the test and control net difference rates

Table 9 shows the level or rate of inconsistent answers between the rooms and bedrooms questions for both the control and test treatment groups on the mail questionnaire. An inconsistent response is defined as when the respondent provides a count of bedrooms that is equal to or greater than the count given for the rooms question. From Table 9, we find that approximately 2.8 percent more answers provided to the rooms and bedrooms questions in the control treatment group are inconsistent than in the test treatment group.

Table 9. Inconsistency Between Rooms and Bedrooms Responses, Control vs. Test (mail only)

	Control (%)	Tests (%)	Difference (%)	Margin of Error (%)	Significant
Reported bedrooms ≥ rooms	6.7	3.8	-2.8	± 0.7	Yes

As part of the test version of the CAPI instrument, a follow up question was asked after the rooms question asking whether the respondent excluded bedrooms in the rooms count. Table 10 shows that 8.9 percent of respondents answered yes to this question. This result provides evidence of the problem where respondents exclude bedrooms when asked to report the number of rooms in their housing unit.

Table 10. Exclusion of Bedrooms in the Rooms Count (CAPI/Test only)

	Test (%)	Margin of Error* (%)
Excluded Bedrooms	8.9	± 1.9

*Does not represent the margin of error of the difference between control and test percent estimates

Changing from a closed- to an open-ended layout and incorporating other changes to the rooms and bedrooms questions reduced underreporting of rooms, increased reporting of "0" bedrooms, and improved consistency between rooms and bedrooms while maintaining the item response rate. The systematic response error for both questions was also maintained or reduced for the test panel with the exception of the "9 or more" rooms category for the rooms question. Based on these results, the test version performed better than the control version. We suspect that the better performance of the test version was mainly due to the question wording and instruction changes. The subject matter experts viewed these content changes as more critical than the response format change.

3.2 Vehicles

Besides changing the vehicles question from a closed-ended to an open-ended format, the additional changes were minimal. The term "SUVs" was added and an instruction to exclude motorcycles and other recreational vehicles was added.

For the vehicles question, nonresponse was defined as no check box checked (control), no entry given in the write-in field (test), or an illegible entry given in the write-in field. Table 11 shows that the item nonresponse rate for vehicles is slightly higher in the test version when compared with the control at the national level. Similar differences are evidenced in the high and low response areas. Note that the rates for both control and test are less than 2.5 percent.

Table 11. Item Nonresponse Rates for Number of Vehicles Question

Strata	Control (%)	Test (%)	Difference (%)	Margin of Error (%)	Significant
National	1.1	1.9	0.8	± 0.4	Yes
High Response Area	1.1	1.8	0.7	± 0.5	Yes
Low Response Area	1.3	2.4	1.2	± 0.4	Yes

The chi-square test in Table 12 indicates no significant differences in the distribution rates for number of vehicles available between the test and control versions at the national level. Further analysis show this is also the case for high and low response areas (tables not shown).

Table 12. Household Number of Vehicles Distribution – National

Number of Vehicles	Control (%)	Test (%)	Difference (%)	Margin of Error (%)	Significant
None	8.3	7.7	-0.6	$\pm$ 0.8	No
1	32.7	33.1	0.4	$\pm$ 1.7	No
2	39.3	39.3	0.0	$\pm$ 1.8	No
3	14.0	14.5	0.4	$\pm$ 1.3	No
4	4.2	3.9	-0.3	$\pm$ 0.7	No
5	1.0	0.9	0.0	$\pm$ 0.3	No
6 or more	0.4	0.5	0.1	$\pm$ 0.3	No
Total	100.0	100.0			

χ^2 = 2.9 with 6 degrees of freedom, not significant at the 10 percent level

The empirical results show that the test version had no impact on the distribution of the number of vehicles. However, the test version did not maintain nor reduce the item nonresponse rate. Based on these results, we conclude that the test version did not perform any better than the control version.

3.3 Property Value

The property value question was answered only by those respondents who reported the household as "owner-occupied." Therefore, we restrict our analysis to those housing units that were reported as being owner-occupied.

For the property value question, nonresponse was defined as no check box checked (control), no entry given in the write-in field (test), or an illegible entry given in the write-in field. Table 13 shows that when compared to the control version, the test version led to a significant increase (4.4 percent) in the level of missing data. We observe a similar result within the high and low response areas. This may provide support for the cognitive test findings where participants expressed that giving an exact property value in the open version was more difficult and burdensome than giving a range for their property value in the closed version. The mail test version asked the respondent to provide the property value "amount – Dollars." Had the test version required a less "accurate" response, we suspect that the respondents to the test version may have had less difficulty providing an estimate of their property value.

Table 13. Property Value Item Nonresponse Rates, Control vs. Test

Strata	Control (%)	Test (%)	Difference (%)	Margin of Error (%)	Significant
National	8.1%	12.4%	4.4%	$\pm$ 1.5%	Yes
HRA	7.4%	11.9%	4.5%	$\pm$ 1.8%	Yes
LRA	11.9%	15.1%	3.3%	$\pm$ 2.2%	Yes

Data from Table 14 show that the test version did not significantly change the median property value at the national level and within high response areas for owner-occupied housing units. However, we observe that the test version significantly reduced the median property value for low response areas. Note that the medians for both the control and test versions were calculated in the same manner as the rooms and bedrooms medians (this required recoding the test version data into the control version's property value categories).

Table 14. Median Property Value, Control vs. Test

Strata	Control ($)	Test ($)	Difference ($)	Margin of Error ($)	Significant
National	$184,979	$174,930	-$10,049	$\pm$ $12,106	No
HRA	$190,211	$181,490	-$8,720	$\pm$ $14,034	No
LRA	$153,557	$138,423	-$15,134	$\pm$ $10,846	Yes

Table 15 shows the property value distributions for the control and test treatment groups. Using a chi-square test, we find that the overall property value distribution differs by question version. Reviewing the individual category comparisons, we observe a large significant difference for the "less than $30,000" category. We hypothesized that this result may suggest that some respondents may be reporting property values in thousands in the open-ended question version. For example, a respondent who indicated "$58" may really be estimating his or her property at "$58K" or $58,000. Performing ad-hoc analysis of the reported annual real-estate tax data reported under the mortgage series for the "less than $30,000" universe, we found that 52 percent of these cases reported annual real-estate taxes of $1,000 or more. This result suggests that a large portion of the "less than $30,000" universe contain reported property values where the zeroes have been truncated. A solution to this problem may be to edit these data using information from established state value-to-tax rates. Editing the test version data using the annual real-estate tax data may help reduce the differences between the test and control distributions. However, with the other unexplained distributional differences coupled with the increased levels of missing data, we suspect that changing from a closed- to an open-ended format will lead to a "break-in-series" for the property value question. A "break-in-series" means that the ACS data collected on property value in series or over time will deviate or break from the traditional pattern observed over time due to changes made to the property value question.

Table 15. Property Value Percent Distribution Rates for Owner Occupied Units, Control vs. Test

Value	Control (%)	Test (%)	Difference (%)	Margin of Error (%)	Significant
Less than $30,000	4.3%	7.0%	2.7%	$\pm$ 1.1%	Yes
$30,000 to $39,999	1.8%	1.6%	-0.2%	$\pm$ 0.6%	No
$40,000 to $49,999	2.4%	1.9%	-0.5%	$\pm$ 0.9%	No
$50,000 to $59,999	1.7%	2.1%	0.5%	$\pm$ 0.6%	No
$60,000 to $69,999	2.7%	2.9%	0.1%	$\pm$ 0.8%	No
$70,000 to $79,999	3.9%	3.3%	-0.6%	$\pm$ 0.8%	No
$80,000 to $89,999	3.8%	4.0%	0.2%	$\pm$ 1.0%	No
$90,000 to $99,999	3.7%	3.0%	-0.8%	$\pm$ 0.7%	Yes
$100,000 to $124,999	6.8%	8.7%	1.9%	$\pm$ 1.1%	Yes
$125,000 to $149,999	8.3%	7.4%	-0.8%	$\pm$ 1.2%	No
$150,000 to $174,999	8.2%	8.1%	0.0%	$\pm$ 1.3%	No
$175,000 to $199,999	6.0%	5.8%	-0.2%	$\pm$ 1.1%	No
$200,000 to $249,999	9.5%	9.0%	-0.5%	$\pm$ 1.2%	No
$250,000 to $299,999	7.0%	7.2%	0.2%	$\pm$ 1.0%	No
$300,000 to $399,999	10.3%	9.1%	-1.2%	$\pm$ 1.3%	No
$400,000 to $499,999	6.2%	5.6%	-0.6%	$\pm$ 1.0%	No
$500,000 to $749,999	8.1%	7.9%	-0.2%	$\pm$ 1.2%	No
$750,000 to $999,999	3.2%	3.0%	-0.2%	$\pm$ 0.6%	No
$1,000,000 or more	2.3%	2.4%	0.2%	$\pm$ 0.5%	No
Total	100.0%	100.0%			

$\chi^2 = 37.1$ with 18 degrees of freedom, significant at the 10.0 percent level

The index of inconsistency (IOI) is the percentage of the total variance that is due to simple response variance for the given response category, and it is a measure of reliability or consistency. IOI values of less than 20 percent indicate high reliability, 20 to 50 percent indicate a moderate level of reliability, and over 50 percent indicate low reliability.

The L-fold index of inconsistency is a weighted average of the individual indexes computed for each response category. This gives an overall measure of reliability for a given question.

Table 16 shows the indexes of inconsistency for the individual property value categories by control and test treatment. In addition, the L-fold IOI is given by treatment. Reviewing the individual categories, we observe a large difference in the IOI between the control and test for the "less than $30,000" category. The significant reduction in reliability for the test version is likely related to the problem we highlighted earlier in Table 15 where respondents are reporting their values in thousands in the write-in field. Reviewing the L-fold indexes of inconsistency, we observe that both control (40.9 percent) and test (36.9 percent) are in the moderate range, but the test version is significantly lower. Therefore, we conclude that the test version of the property value question is more reliable than the control version.

Table 16. Property Value Content Followup Comparison Statistics – Index of Inconsistency, Control vs. Test (Mail Only)

Property Value	Control vs CFU (%)	Test vs CFU (%)	Diff (%)	Marg. Err (%)	Signif
Less than $30,000	19.4%	55.9%	36.5%	+7.4%	Yes
$30,000 - $39,999	54.4%	47.5%	-6.8%	+13.8%	No
$40,000 - $49,999	61.9%	43.9%	-18.0%	+11.5%	Yes
$50,000 - $59,999	52.9%	50.4%	-2.6%	+10.6%	No
$60,000 - $69,999	53.9%	48.2%	-5.7%	+9.5%	No
$70,000 - $79,999	52.7%	48.4%	-4.3%	+9.1%	No
$80,000 - $89,999	51.4%	43.7%	-7.7%	+8.3%	No
$90,000 - $99,999	51.4%	49.1%	-2.3%	+8.7%	No
$100,000 - $124,999	45.2%	35.2%	-10.0%	+4.9%	Yes
$125,000 - $149,999	43.7%	34.5%	-9.2%	+5.5%	Yes
$150,000 - $174,999	43.7%	38.5%	-5.2%	+5.5%	No
$175,000 - $199,999	50.1%	38.6%	-11.5%	+5.7%	Yes
$200,000 - $249,999	40.4%	32.5%	-7.8%	+4.4%	Yes
$250,000 - $299,999	50.2%	37.2%	-13.1%	+5.2%	Yes
$300,000 - $399,999	33.5%	31.1%	-2.4%	+4.9%	No
$400,000 - $499,999	34.0%	34.1%	0.1%	+5.6%	No
$500,000 - $749,999	22.1%	24.0%	1.9%	+4.0%	No
$750,000 - $999,999	28.9%	32.8%	3.9%	+7.0%	No
$1,000,000 or more	17.6%	26.5%	8.8%	+6.8%	Yes
L-fold	40.9%	36.9%	-4.0	+1.8%	Yes

The median values are similar for the ACS test and control versions, and the reliability for the test version is significantly higher than the control version overall. Incorporating the test version changes may come at the cost of higher non-response rates to the property value question. However, it will facilitate the collection of more precise data and thus serve the needs of HUD, the major Federal user of housing statistics. Note that a possible solution to alleviating some of the item nonresponse error associated with the open-ended property value question would be to ask the respondent for a less exact value. For example, asking for a value to the nearest $1000 amount as opposed to the nearest dollar amount.

4. Conclusion

Across the question topics tested in the content test we measured the incidence of missing data using item nonresponse rates and compared the resulting rates between the open- and closed-ended formats. In general, we found that the open-ended version of the questions for the rooms, bedrooms, and vehicles questions either maintained the current level of missing data or marginally increased it. For the property value question, we observed a significant increase in the incidence of missing data for the open-ended version. The increase in INR for property value was likely due to the increased respondent burden of requiring the respondent to give an exact property value versus identifying a value range.

In addition to determining the impact on item missing data, we studied the impact the open- and closed-ended formats had on the response distributions. With the exception of the vehicles question, we found that the question versions using the open-ended layout affected the response distribution. Even though we were unable to isolate the response distribution effects of the open-ended format for the rooms and bedrooms questions, we suspect that the differences we observed in the response distributions were likely due to the other changes coupled with the open format (i.e., providing a rooms definition, re-wording the bedrooms question to link it to the rooms question, providing an efficiency apartment instruction, etc.). However, for the property value question, we were able to attribute some of the change in the property value response distribution directly to changing from a closed to an open-ended response option. By examining the annual real-estate tax data we discovered that a large portion of the open-ended responses for the "30,000 or less" category were responses where the respondent wrote in a property value with the zeros truncated. While this respondent behavior may explain some of the differences in the property value distributions between the control and test treatment groups, some of the other distributional differences may be driven by the higher item nonresponse produced by the open-ended version.

Including the rooms, bedrooms, and property value questions as part of the CFU study, we were able to study what effects the open-ended version of these questions had on systematic response error and reliability. For the rooms and bedrooms questions we found that the open-ended version coupled with the other changes made to the rooms and bedrooms questions maintained or reduced the amount of systematic response error produced by the closed-ended version for most categories. As for the property value question, we found that the open-ended version produced higher levels of reliability comparable to those produced by the closed-ended version.

Based on the ACS Content Test results for these select housing items, the open-ended format appears to be a feasible alternative to the closed-ended format for collecting data for the rooms, bedrooms, and vehicles items at the national level. As for the property value question, we cannot as easily conclude that the open-ended version is a worthy alternative to the closed-ended format because of the increased item nonresponse and the respondent behavior of truncating property values. However, these negative aspects of the open-ended version may be worth the cost to data users given that the data are more reliable and can easily be adjusted for inflation. Note that the OMB decision for final content on the 2008 ACS questionnaire approved the use of the open-ended versions of the property value, rooms, and bedrooms questions. The vehicles question will remain a closed-ended question.

References

Asiala M. and Navarro A. (2006). "Experimental Design for the 2006 American Community Survey Content Test," American Statistical Association 2006 Proceedings of the Section on Survey Research Methods [CD-ROM].

Bradburn, N.M., Sudman, S., and Associates (1979). *Improving Interview Method and Questionnaire Design*, San Francisco, Jossey-Bass.

Converse, J. and Presser, S. (1986). *Survey Questions: Handcrafting the Standardized Questionnaire*, Newbury Park, Sage Publications Inc.

Kerwin, Heltemes, Franklin, Nelson, and Popovic (2005) "Cognitive Testing of Proposed Items on Housing for the American Community Survey," WESTAT, Rockville, MD.

Rao, J. and Scott, A. (1981). "The Analysis of Categorical Data from Complex Sample Surveys: Chi-Squared Tests for Goodness of Fit and Independence in Two-Way Tables," Journal of the American Statistical Association, vol. 76, 221-230.

Rao, J. and Scott, A. (1984). "On Chi-Squared Tests for Multiway Contingency Tables with Cell Proportions Estimated from Survey Data. Annals of Statistics, vol. 12, 46-60.

Schuman, H. and Presser, S. (1981). *Questions and Answers in Attitude Surveys*, New York, Academic Press, 1981.

Singer P. and Ennis S. (2003), "Census 2000 Content Reinterview Survey: Accuracy of Data for Selected Population and Housing Characteristics as Measured by Reinterview", Census 2000 Evaluation B.5, U.S. Census Bureau, Washington.

Sudman, S., Bradburn, N., and Schwarz, N. (1996). *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*, San Francisco, Jossey-Bass.

Appendix

Figure 1. Rooms Control and Test Questions

How many rooms are in this house, apartment, or mobile home? *Do NOT count bathrooms, porches, balconies, foyers, halls, or half-rooms.*

- ☐ 1 room
- ☐ 2 rooms
- ☐ 3 rooms
- ☐ 4 rooms
- ☐ 5 rooms
- ☐ 6 rooms
- ☐ 7 rooms
- ☐ 8 rooms
- ☐ 9 or more rooms

How many separate rooms are in this house, apartment, or mobile home? *Rooms must be separated by built-in archways or walls that extend out at least 6 inches and go from floor to ceiling.*
- *INCLUDE bedrooms, kitchens, etc.*
- *EXCLUDE bathrooms, porches, balconies, foyers, halls, or unfinished basements.*

Number of rooms

| |

Figure 2. Bedrooms Control and Test Questions

How many bedrooms are in this house, apartment, or mobile home; that is, how many bedrooms would you list if this house, apartment, or mobile home were on the market for sale or rent?

- ☐ No bedroom
- ☐ 1 bedroom
- ☐ 2 bedrooms
- ☐ 3 bedrooms
- ☐ 4 bedrooms
- ☐ 5 or more bedrooms

How many of these rooms are bedrooms? *Count as bedrooms those rooms you would list if this house, apartment, or mobile home were for sale or rent. If this is an efficiency/studio apartment, print "0".*
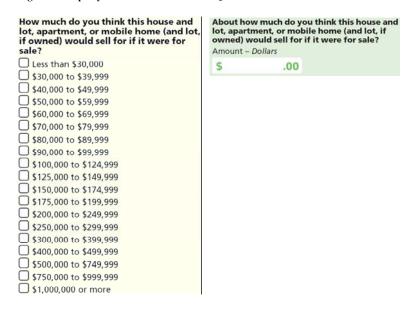
Number of bedrooms

| |

Figure 3. Vehicles Control and Test Questions

How many automobiles, vans, and trucks of one-ton capacity or less are kept at home for use by members of this household?

- ☐ None
- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5
- ☐ 6 or more

How many automobiles, vans, SUVs, and trucks of one-ton capacity or less are kept at home for use by members of this household? *Do not include motorcycles or other recreational vehicles.*

Number of vehicles

| |

Figure 4. Property Value Control and Test Questions

How much do you think this house and lot, apartment, or mobile home (and lot, if owned) would sell for if it were for sale?

- ☐ Less than $30,000
- ☐ $30,000 to $39,999
- ☐ $40,000 to $49,999
- ☐ $50,000 to $59,999
- ☐ $60,000 to $69,999
- ☐ $70,000 to $79,999
- ☐ $80,000 to $89,999
- ☐ $90,000 to $99,999
- ☐ $100,000 to $124,999
- ☐ $125,000 to $149,999
- ☐ $150,000 to $174,999
- ☐ $175,000 to $199,999
- ☐ $200,000 to $249,999
- ☐ $250,000 to $299,999
- ☐ $300,000 to $399,999
- ☐ $400,000 to $499,999
- ☐ $500,000 to $749,999
- ☐ $750,000 to $999,999
- ☐ $1,000,000 or more

About how much do you think this house and lot, apartment, or mobile home (and lot, if owned) would sell for if it were for sale?

Amount – *Dollars*

$ _____ .00