Monitoring, Experience and Performance of CATI Interviewers in the American Community Survey[1]

Dale Garrett, Andre Williams and Rita Petroni
Dale Garrett, U.S. Bureau of the Census, Washington, D.C. 20233
b.dale.garrett@census.gov

Key words: feedback, quality control

Abstract

The American Community Survey (ACS) collects important socioeconomic data continuously in three month cycles. ACS mails out questionnaires during month one, then follows up nonrespondents with computer assisted telephone interviewing (CATI) in the second month and personal interviewing in the third month. Quality control for CATI is performed by monitoring interviewers and providing them with regular feedback on their performance. Interviewers do not know exactly when and how often they are monitored. This paper addresses the consistency of interview quality by examining the ratings assigned to monitored interviews and the elements (parts) of the interviews. This paper reports distributions of interviewer performance for data available from January 1, 2001 through March 31, 2003. This paper also discusses changes in interviewer performance during this period and the possible link to increased interviewer experience.

1. Introduction

The American Community Survey (ACS) has been developed by the U. S. Census Bureau to provide the data traditionally collected by the decennial census long form sample. This program has been tested and phased in over several years. When the ACS is fully implemented, it will replace the census long form sample. The ACS will use a five-year rolling sample to provide small area estimates for demographic, social, economic and housing characteristics.

The ACS collects data through three modes in a three month cycle. During the first month, the ACS collects data through mailout/mailback. During the second month, sample addresses which were mail nonrespondents and for which the ACS has telephone numbers are sent to one of three telephone centers for computer assisted telephone interviewing (CATI). During the third month, any cases which could not be interviewed by mail or by CATI, are subsampled. Selected cases are sent to the field for computer assisted personal interviewing (CAPI). For more information on ACS operations see the "American Community Survey Operations Plan: Release 1" (2003).

This paper presents data describing the monitoring of interviewers during the CATI phase of the ACS. Note that the CATI phase obtains interviews for just under ten percent of the ACS sample. One objective of the CATI monitoring program is to detect and deter falsification. It also provides constructive and positive feedback designed to improve interviewing performance. This evaluation uses archived data from CATI monitored interviews to examine the consistency of interviewer performance during the period from January 1, 2001 through March 31, 2003. It examines changes in interviewer performance during this period and the possible link between interviewer experience and improved interviewer performance.

2. Background

ACS supervisors at three call centers monitor the telephone interviews to evaluate job performance, and to provide feedback to interviewers to improve data quality. The supervisors often monitored two or more interviewers during the same period, during which they could have switched between interviewers several times. In this paper, a **monitored interview** is defined as the activity of a single interviewer during a period of time in which the supervisor listened to portions of at least one interview.

During the monitored interview, a supervisor rates up to **eight elements** which include: Introduction, Manner/Voice, Reading Skills, Probing Skills, Response Entries, Difficult Situations, Survey Concepts, and Session Notes. The five rating levels assigned to elements are:

1 - needs improvement,
2 - marginal,
3 - fully successful,
4 - commendable, and
5 - outstanding.

[1]This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed on operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

The rating assigned to each monitored interview is the average of the element ratings.

Interviewers at the three CATI facilities are classified as **initial, systematic** and **special needs.** An **initial** interviewer is either a new hire or an experienced interviewer from another survey being cross-trained on the ACS. Each new hire must achieve a minimum rating of fully successful on **all elements** on each of three consecutive monitored interviews before being classified as a **systematic** interviewer. A cross-trained interviewer needs to have only one completely successful interview before being classified as a **systematic** interviewer. An interviewer remains in systematic status unless he/she has three consecutive monitored interviews rated below fully successful, in which case his/her status drops to **special needs**. An interviewer in special needs status must have three consecutive monitored interviews fully successful on all elements to return to systematic status. Staff members in special needs status are provided feedback on what caused the deficiency and how to improve.

Supervisors monitor a minimum of 2.5 percent of the logged-in hours of each systematic interviewer and 5 percent of the logged-in hours of each initial and special needs interviewer. They often monitor more than the specified percents. Feedback is provided to interviewers for each rated interview. About 2 out of 3 monitored interviews were rated during January 2001 through March 2003. Feedback forms are printed out at the conclusion of each rated interview. Supervisors review each feedback form with the interviewer and have the form signed by the interviewer to indicate that they have received the completed form and discussed it as needed with the supervisor. Feedback must be specified as either 'praise' or 'improve' by element. To improve consistency in ratings, monthly sessions are conducted in which supervisors from all three call centers are trained to calibrate ratings.

3. Methods

Research questions included:

- What proportion of interview elements, interviews and CATI interviewers had ratings below fully successful? Did the ratings indicate a high level of success?
- Did the ratings change between periods (January 1, 2001 through December 31, 2001 compared to January 1, 2002 through March 31, 2003)?
- Did the ACS CATI interviewers in the second period have more experience than in 2001?
- Did the interviewers have more consistent ratings as the level of experience increased?

Rating distributions were examined for each element of a monitored interview (Tables 2, 3 and 4), and for each monitored interview (Table 7).

For Tables 6 and 7 a simple weighting scheme was applied. Since the initial and special needs interviewers were monitored twice as frequently as the systematic interviewers, their monitored interviews were given one half the weight assigned to systematic interviewers.

The year 2001 data were compared to the data from the fifteen month period: January 1, 2002 through March 31, 2003. The CATI monitoring data files available at the time of this research included twelve months of data from 2001, nine months of data from 2002, and three months from 2003. During 2002, there were no ACS CATI interviews during July and August because of reductions in the ACS sample. Also, all monitor data for March 2002 were not accessible in the electronic files available to the researchers. Although paper records of monitor feedback data were available, it was not feasible to recreate the March 2002 file. We added the first three months from 2003 to give us 12 months to compare to 2001. We refer to the period from January 2002 through March 2003, as **2002+**.

Statements of comparisons or implied comparisons in this paper, were tested for statistical significance using alpha equal to 0.1.

4. Results

4.1 CATI Interviewers by Type

The first row of Table 1, beginning with the first column shows that there were 402 CATI interviewers who worked on the ACS sometime during January 1, 2001 through March 31, 2003. Of these, 355 spent some of the time as systematic interviewers, 243 spent some of the time as initial interviewers and 31 spent some of the time as special needs interviewers. Rows 2 and 3 further breakdown these categories by the two time periods, 2001 and 2002 + respectively. This table shows that the CATI workforce required fewer initial interviewers during 2002+ because more of the systematic interviewers were retained.

Table 1. Counts of CATI Interviewers for January 1, 2001 through March 31, 2003

		Type of Interviewer		
	Total ** Interviewers	Systematic Interviewers	Initial Interviewers	Special Needs Interviewers
Total Period **	402	355	243	31
2001	332	274	188	23
2002 + *	267	250	70	14

* The fifteen month period from 01/01/2002 to 03/31/2003 contained twelve months of ACS CATI data and is referred to as 2002 + .

** Rows and columns do not sum to totals because an interviewer may spend time as initial, systematic and special needs within the same year and may work both years.

4.2 Distribution of Ratings by Element

This section provides information to understand the big picture of element ratings.

Tables 2 and 3 provide distributions of ratings assigned to one or more elements of interviews based on 43,653 instances of rated elements during January 1, 2001 to March 31, 2003. Table 2 shows the distribution of element ratings for all elements while Table 3 provides the distribution by individual element.

From Table 2 we see that 91.6 percent of total element ratings were fully successful. The fully successful rating is considered the standard. To receive a rating of commendable or outstanding for an element, an interviewer must do something above and beyond the ordinary, such as deal successfully with a difficult respondent.

Adding the three columns on the right of Table 2, we find that about 96.9 percent were fully successful or better, leaving only 3.1 percent of elements with poor ratings.

Table 3 provides more details of element ratings. For each rated interview, a subset of elements is rated depending on the situation and the portions of the interview heard by the monitor. The penultimate column of Table 3 shows the number of records for which the element was rated. For example, on the second row (Introduction), this column indicates that 1,948 records of monitored interviews were assigned a rating for the introductions. That is only 20.4 percent of the 9,542 rated records as shown in the last column. When a monitor chooses an interview to rate, chances are the rated interview was already in progress, so they missed the introduction most of the time. However, the introduction was rarely a trouble spot in the interview as shown by the Fully Successful column, which indicates that 98.2 percent of rated introductions were assigned a fully successful rating. Beginning in 2004, the monitor program will display the progress of the interviewer on a call, so that the monitor has a better chance to start listening right after contact is made. Other elements which were rarely rated include difficult situations and notes. Difficult situations were rated 10.4 percent of the time. Difficult situations may occur when a respondent is uncooperative or speaks very little English or when something occurs that makes the interview difficult to complete. The notes section was rated 0.5 percent of the time. The notes section is normally used only to record case numbers and is only rated when special notes are required.

Reading Skills had the most low ratings. About 6.2 percent of the ratings of Reading Skills were below the standard. A review of monitor comments indicates that these low ratings were caused by skipping parts of questions or skipping entire questions. For example, the race question includes 15 check boxes for different race categories. In some cases an impatient respondent answered before the interviewer had read the entire question, and the interviewer did not read the rest of the race categories. The ACS allows multiple responses to the race question, so it is important that the respondents hear all of the race categories even though they have already answered the question. Interviewers improved in reading skills during this period. The percent below fully successful decreased from 8.2 percent in 2001 to 4.2 percent in 2002+.

Table 2. Ratings Distribution for 43,653 Rated Elements

Unweighted

	Ratings				
	Needs Improvement	Marginal	Fully Successful	Commendable	Outstanding
Proportion of Total Element Ratings	1.4%	1.7%	91.6%	4.6%	0.7%

Table 3. Ratings Distribution by Individual Element

Unweighted

	Ratings					Monitored Interviews with Rating Assigned	
Element	Needs Improvement	Marginal	Fully Successful	Commendable	Outstanding	#	% of 9,542
Difficult situations	0.9%	1.1%	58.3%	29.8%	9.8%	989	10.4%
Introduction	0.5%	0.9%	98.2%	0.4%	0.0%	1,948	20.4%
Manner/ voice	0.4%	1.0%	91.3%	6.2%	1.0%	9,290	97.4%
Reading skills	3.5%	2.7%	89.3%	4.1%	0.4%	9,005	94.4%
Probing skills	0.9%	2.0%	92.0%	4.9%	0.3%	5,513	57.8%
Response entries	1.1%	1.5%	94.6%	2.8%	0.0%	9,369	98.2%
Survey concepts	1.0%	1.6%	93.7%	2.7%	0.9%	7,487	78.5%
Notes	7.7%	0.0%	50.0%	42.3%	0.0%	52	0.5%
					Total	43,653	elements

4.3 Changes in Total Element Ratings

The changes in element ratings between 2001 and 2002+ have been positive. As shown in row two of Table 4, the proportion of element ratings which were at least fully successful increased from 95.9 percent in 2001 to 98.0 percent in 2002+. This indicates that interviewers tend to do their job consistently, usually scoring a fully successful or higher for each element and usually averaging a fully successful or better for each monitored session.

Table 4. The Proportion of Element Ratings Below Fully Successful vs At Least Fully Successful

Unweighted

	Year	
	2001	2002 +
Below Fully Successful	4.1%	2.0%
At Least Fully Successful	95.9%	98.0%
Number of Elements	22,078	21,575

4.4 Why the Fully Successful Rating Is Important

Monitoring seeks to identify performance issues. Poor performance in interview elements may affect response rates, quality of data, and/or public perception of the Census Bureau. Successful performance closes the door to some types of nonsampling error. For example, during the introduction, monitors listen to make sure that interviewers verify the address, and the eligibility of the respondent. If the interviewer does not verify these items, we may be interviewing someone who is not in sample, thus introducing nonsampling error.

The monitor also listens to ensure that interviewers read questions in their entirety as worded, follow the correct skip patterns and accurately record the response. The monitor also listens to make sure that the interviewer probes for an accurate response when the respondent is vague or contradictory. The fully successful rating is important here since it helps ensure that we are collecting quality data. The monitor also listens to make sure that the interviewer attempts to convert a reluctant respondent. Reducing nonresponse also may reduce nonsampling error and potential bias.

Table 5 shows a summary of types of monitor comments by element for cases when the elements were rated below fully successful.

Table 5. Reasons That Monitors Rated an Element Below Fully Successful

Element	Reasons That Elements Were Rated Below Fully Successful
Difficult Situations	- no attempt made to convert a reluctant respondent - did not use available resources to answer a reluctant respondent's questions - provided insufficient answers to respondents questions
Introduction	- failed to inform respondent that interview may be monitored for quality assurance - interviewer did not identify himself/herself to the respondent - interviewer did not identify Census Bureau and/or survey - conducted an interview with an ineligible respondent
Manner/Voice	- made unnecessary personal comments - spoke negatively against survey or questions on survey - talking too fast where respondent cannot understand what is being asked
Reading Skills	- did not read some questions in their entirety - skipped reading some questions
Probing Skills	- answered questions for the respondent rather than probing for the correct response - accepted unclear responses rather than probing for clear ones
Response Entries	- entered responses prior to the respondent answering the questions - entered a response not given by the respondent - entered a response without asking the questions to the respondent
Survey Concepts	- interviewer did not identify himself/herself when changing to a new respondent - interviewer provided answers for the respondent - interviewer did not verify address information or telephone information

4.5 Interview Quality Increased with Interviewer Experience and Other Factors

The proportion of monitored interviews which were rated at least fully successful (from 3 up to 5) is one measure of the quality of the CATI interview. This measure improved from 89.8 percent of monitored interviews in 2001 to 94.7 percent in 2002+. This section looks at the possible relationship between interviewer experience and the increase in the number of interviews with ratings equal to or above fully successful.

During the CATI activity in 2002+, there were more interviews conducted by experienced interviewers than in 2001. This increase in interviewer experience helped increase the proportion of monitored interviews which were at least fully successful.

The fifth row of Table 6 (Performed by More Experienced Interviewers) shows that only 61.6 percent of the monitored interviews were conducted by more experienced interviewers in 2001 as compared to 79.1 percent in 2002+.

To understand the relationship between experience and ratings, we categorized monitored interviews by the level of ACS experience of the interviewer. Any interviewer in systematic status by the end of November 2000, was labeled as an early experience interviewer. For other interviewers, the level of experience was defined by counting the number of ACS monitored interviews after the interviewer became a systematic interviewer. If the interviewer spent some time as a special needs interviewer, we counted each interview during special needs status as one half of a systematic interview since the special needs interviewers are monitored twice as often.

Table 6 provides a snapshot of the change in experience between 2001 and 2002+. It shows the percent of rated monitored interviews at each level of experience by year. Percents are weighted so they represent an estimate of the proportion of ACS interviews. The first row under Less Experienced (Initial Interviewers) in table 6 indicates that the percent of ACS CATI interviews conducted by initial interviewers decreased from 5.6 percent in 2001 to 2.1 percent in 2002+.

Table 7 shows that the quality of the interview improved as interviewer experience increased. The quality of the interview here is measured by the proportion of interview ratings which were at least fully successful. First look at the columns separately. Examine the 2001 column, comparing the third, fourth and fifth rows. We find that this proportion was only 72.8 percent for initial interviewers, but increased to 88.2 percent for the new systematic interviewers (with 1 to 20 systematic monitored interviews), and averaged 92.2 percent for the More Experienced systematic interviewers.

The 2002 column also shows an increase related to experience. Again moving from the third to fifth rows, we find the proportion increased from 80.3 percent for the initials up to about 92.0 percent for the new systematic interviewers and 95.8 percent for the More Experienced systematic interviewers.

The last row of Table 7 shows a similar story, but we must read the change in experience occurring within the row instead of within a single column. The Early experience systematic interviewers had on the average twelve months more experience during 2002+ than they did in 2001. The change from 91.0 to 95.9 percent at least fully successful is significant and supports the claim that the proportion fully successful increased as experience increased.

However, looking at the second, third and fourth rows we find what appears to be a contradiction to our hypothesis that the improvement in the proportion fully successful is due to experience. The Less Experienced row shows that less experienced interviewers in 2001 had only 85.9 percent at least fully successful, compared to 90.9 percent in 2002+. This change is also statistically significant. We interpret this to mean that the new ACS CATI interviewers in 2002+ adapted more quickly than their counterparts in 2001. Reasons for this change between time periods may be related to the fact that during 2002+ the ACS had fewer initials (70 compared to 188 in 2001). Fewer new interviewers could mean several things, any of which could lead to this improvement: a more selective hiring process, more personal attention from supervisors during training and the early stages of their work experience on the ACS, and a work environment with a higher percentage of experienced staff doing quality work.

Table 6. Percent of Rated Monitored Interviews by Level of Interviewer Experience
Weighted

Level of Interviewer Experience	2001	2002+
Total Rated Interviews	100.0%	100.0%
Performed by Less Experienced Interviewers	38.4%	20.9%
• Initial Interviewers	5.6%	2.1%
• with 1 to 20 systematic monitored interviews *	32.8%	18.8%
Performed by More Experienced Interviewers	61.6%	79.1%
• with 21 to 50 systematic monitored interviews *	16.2%	21.8%
• with 50 or more systematic monitored interviews *	1.4%	23.7%
• Early experience systematic interviewers * (achieved systematic status by November 30, 2000)	44.0%	33.6%
Number of monitored interviews which had a rating assigned.	4,827	4,715

Table 7. Percent of Interviews with an Interview Rating of At Least Fully Successful by Level of Interviewer Experience
Weighted**

	2001		2002+	
Level of Interviewer Experience	Percent	Base	Percent	Base
Total	89.8%	4,481	94.7%	4,574
Less Experienced	85.9%	1,722	90.9%	958
• Initial Interviewers	72.8%	504**	80.3%	193**
• 1 to 20 systematic monitored interviews*	88.2%	1,470	92.0%	861
More Experienced	92.2%	2,759	95.8%	3,616
• 21 to 50 systematic monitored interviews*	95.0%	725	94.2%	996
• 50 or more systematic monitored interviews*	95.4%	65	97.0%	1,084
• Early experience systematic interviewers * (achieved systematic status by November 30, 2000)	91.0%	1,969	95.9%	1,536

*The experience of interviewers is shown by the number of monitored interviews after achieving systematic status. Interviewers who achieved systematic status by November 30, 2000 were excluded from the tallies which counted monitored interviews and included on the last row, because we didn't know exactly when in 2000 they became systematic interviewers, and the available files did not provide sufficient data before November 2000.

**The base shown for the initials is unweighted so the reader can see the number of records used. For the other cells, the number shown is close, but usually less than the actual number of records since the records of special needs interviewers and initial interviewers were given weights of 0.5. The records of systematic interviewers were given weights of 1.0.

5. Conclusions

The CATI monitoring program provides an incentive for interviewers to be consistently fully successful in all elements of the interview. Both interview ratings and element ratings have improved between 2001 and 2002+. Much of the improvement in quality is due to the increase in staff experience. Other factors may include the stability of a work environment that brings in new interviewers in smaller numbers.

6. References

Garrett, D. and Williams A. "Evaluation of Monitoring of Computer Assisted Telephone Interviewing for the American Community Survey for January 2001 through March 2003," internal Census Bureau memorandum from Killion to Singh, June 15, 2004.

U.S. Census Bureau, "*Monitoring Performance*", internal memorandum from Woods to Berger, Wright and Benjamin, January 17, 2001.

U.S. Census Bureau, "American Community Survey Operations Plan: Release 1", March 2003.

U.S. Census Bureau, "CATI Monitoring Windows 95 User Guide", undated internal memorandum used in CATI telephone centers during 2001-2002.