USING KEYSTROKE FILES TO ASSESS RESPONDENT DIFFICULTIES WITH AN AUDIO-CASI INSTRUMENT

Rachel A. Caspar, Research Triangle Institute and Mick P. Couper, Joint Program in Survey Methodology
Rachel A. Caspar, RTI, P.O. Box 12194, Research Triangle Park, NC 27709-2194

Keywords: National Household Survey on Drug Abuse, computer-assisted interviewing

1. Introduction

As more and more field surveys are conducted using computer-assisted interviewing (CAI) techniques, new mechanisms for monitoring how interviewers work with these systems have been developed (see for example, Couper, et al., 1994). One monitoring method which can be extremely useful is the keystroke file. A keystroke file maintains a record of every key pressed by the interviewer as he/she moves through the instrument. Analyses of keystroke files may identify where interviewers are having difficulty with the system or where they are not following study procedures. However, to date, keystroke files have not been analyzed with regard to how respondents navigate through a questionnaire designed for computer-assisted self-interviewing (CASI). This paper will describe such an investigation.

The Substance Abuse and Mental Health Services Administration (SAMHSA), sponsor of the National Household Survey on Drug Abuse (NHSDA) has contracted with RTI to determine the feasibility of converting the NHSDA from a paper-and-pencil instrument to a CAPI/Audio-CASI instrument. As part of this conversion work, a keystroke file was incorporated into a small CAI field test (n=400) to gain an understanding of how respondents maneuvered through the Audio-CASI portion of the interview.

In its traditional paper and pencil format, the NHSDA includes both interviewer-administered questions (where the interviewer reads the questions aloud and records the respondent's answers in a questionnaire booklet) and questions that the respondent completes in a self-administered manner using answer sheets. The questions included on the answer sheets are more sensitive in nature (e.g., recency and frequency of illegal drug use) and this self-report format has been shown to increase the reporting of these sensitive behaviors (Turner, et al., 1992). In developing the CAI NHSDA instrument, those questions included on the answer sheets were programmed for Audio-CASI. Questions that the interviewer reads aloud were programmed for CAPI administration.

The field test included two versions of the CAI NHSDA instrument (as well as the current paper NHSDA for comparison purposes). The first, was designated as the "Mirror Image" or MI version. In this version, every attempt was made to replicate the paper instrument on the computer. Thus, no additional skip logic was programmed beyond that which is included in the paper instrument. In the second version (designated as "Skip"), additional skip routing was included to move respondents out of question modules that do not apply to them. In each CAI instrument, a keystroke file was incorporated to collect every keystroke made by either the respondent or the interviewer. However, the analyses presented below document only our findings from the respondent-administered (or Audio-CASI) portion of the interview.

Our analyses were designed to provide information about the relative ease with which respondents worked with the CAI NHSDA instruments and the quality of the data collected. Specifically, our analyses cover the following areas: 1)Use of function keys and other special keys by respondents, 2)Respondents' ability to provide open-ended responses, 3)Indicators of enhanced privacy for the Audio-CASI interview, and 4)Timing comparisons between actual interview data and "gold standard" interviews.

2. Use of Special Keys

In the Audio-CASI portion of the NHSDA interview, respondents had a number of special keys that they could use as they completed the interview. These keys included:

F2 — Clear Range Error Message
F3 — Don't know
F4 — Refusal
F5 — Lower volume of recorded voice
F6 — Raise volume of recorded voice
F7 — Turn the screen on/off (a toggle switch)
F8 — Turn the sound on/off (a toggle switch)
F9 — Backup one screen at a time
F10 — Replay audio for the screen

In addition, respondents could use the Delete and Backspace keys to edit their answers, as well as the left, right, up, and down arrow keys. Respondents' use of these special keys may indicate difficulty working with the Audio-CASI instrument. Conversely, it could indicate that respondents took advantage of the special

keys to make the task of completing the interview easier and their answers more accurate. Table 1 shows the number of respondents who used any special keys during the self-administered portion of the NHSDA. In these tables, we have collapsed the special keys into four categories to facilitate review of the data.[1]

Data from Table 1 show that many respondents made use of the additional keys available to them. More than two-thirds of the respondents (68%) used an editing key at least once during the Audio-CASI portion of their interview. The rate is somewhat higher for MI respondents than for SKIP respondents (74% vs. 62%). This could be due to the fact that MI respondents were required to answer every question in each core section regardless of whether they reported use of the substance or not. This may have been a more difficult task for respondents and one which caused more respondents to need to review or revise their answers.

Table 1 also shows the percentage of respondents who pressed the Escape key during the Audio-CASI interview. It should be noted that respondents were never instructed to use the Escape key as part of the Audio-CASI tutorial they completed at the beginning of the interview. However, in other software packages, the Escape key is often a means of "getting out of trouble." It is possible that respondents who are more familiar with computers may have tried to use this key more often than other respondents. There is some support for this theory as we found that there were no respondents with less than a high school diploma who used this key.

Youth respondents were especially likely to edit their answers (data not presented here). More than 80 percent of these respondents used an editing key at least once. Youth respondents were also more likely than any other demographic subgroup to use the audio/screen controls. Nearly half (45%) of the youths used an audio/screen control key at least once. It is possible that these respondents feel more comfortable using the special functions provided by the Audio-CASI system because of their increased familiarity with computers. These respondents may also have been more curious to see how all the different functions of the computer operated and thus were simply using these keys to see what would happen.

Table 2 shows the average number of times the special keys were used (among respondents who used the key at all). From this data, we see there is little difference in the number of times the special keys were used by CAI version.

[1]Response Options include the F3 and F4 keys. Audio/Screen Controls include the F5, F6, F7, F8, and F10 keys. Editing keys include the F9, F2, and four arrow keys.

Finally, there were some respondents who used no special keys as they entered their answers in the Audio-CASI interview. This means that these respondents pressed only the appropriate number keys and the enter key for each screen. They never reviewed or revised any of their answers, and never adjusted any of the audio/screen controls. When we reviewed this data, we found that a larger percentage of SKIP respondents used no special keys than did MI respondents (30% vs. 16%). This may be an indication that the SKIP instrument was easier for respondents to work with than was the MI instrument.

3. Open-Ended Responses

The CAI NHSDA instrument included a number of opportunities for respondents to key open-ended text. Eleven of the 20 self-administered sections include at least one opportunity for a respondent to answer with an open-ended response. Across both CAI instruments (MI and SKIP), there were 63 occurrences of a respondent triggering a question that required an open-ended response. These 63 occurrences were attributed to 41 respondents. By reviewing the keystroke files, we can determine whether more respondents actually were routed to the open-ended text screens but then chose to backup to the gate question, revised their answer and went down a different path in order to avoid the keying task. To the extent that this is happening, the Audio-CASI methodology would be reducing the quality of the data collected rather than improving it.

Fortunately, our review of the keystroke files found only three cases where a respondent chose to backup to a previous question and revise his/her answer after being routed to an open-ended text screen. It is impossible to know whether these three respondents were attempting to avoid the keying task or were simply correcting the answer to the previous question based on further thought or after realizing they had mis-keyed their answer. However, given such a small number of respondents made this type of revision, it seems safe to hypothesize that the quality of the data being collected in the open-ended screens is not being seriously affected by the move to computer-assisted data collection.

Although we did not specifically tally the number of occurrences, in our review of the keystroke file we also noticed occasions when a respondent revised an answer such that they were routed to an open-ended text screen.

An additional review of the keystroke files indicates that respondents who did key open-ended text did so with seemingly little difficulty. In most cases, the keystroke file showed that respondents keyed their text in a single pass — that is, they did not need to back up to revise the answer/to fix "typos"/etc. In some cases, however, the keystroke file showed that respondents used the delete

and backspace keys to edit their answer.

Based on these findings, it appears that respondents are capable of keying open-ended text within the Audio-CASI portion of the interview. In addition, respondents do not appear to revise their answers in order to avoid the keying task. However, it should be noted that we cannot determine from the keystroke file whether previous "encounters" with open-ended text screens cause respondents to be more cautious in triggering these screens later in the interview. Nevertheless, given the small number of NHSDA data items collected by requiring the respondent to key open-ended text it seems appropriate to continue to allow respondents to key this text rather than dropping the questions or developing a close-ended questioning format to collect the data.

4. Indications of Enhanced Privacy With Audio-CASI

One of the primary benefits of converting the self-administered portions of the NHSDA to an Audio-CASI format is the ability to improve the privacy of the interview setting. Respondents can read the questions from the computer screen, hear the questions read over the headphones, and enter their answers themselves, thereby reducing the likelihood that the interviewer or other people in the household will know how they answered the questions. This enhanced privacy is theorized to promote more honest reporting of sensitive behaviors by the respondents.

One additional privacy-enhancing feature of the Audio-CASI instrument is the opportunity to turn the screen off. With the screen off, a respondent can still listen to the audio portion of the interview and enter his/her answers into the computer as instructed by the recording. The keystroke file enables us to determine how often respondents chose to turn the screen off and for what proportion of the interview they chose to answer questions in this manner. Overall, only a small percentage of respondents (4.7%) ever turned the screen off. However, looking at a demographic breakdown (data not shown here), we found that 13.4 percent of youth respondents (ages 12-17) turned the screen off at least once. This finding is consistent with the theory that youth respondents may benefit most from the Audio-CASI methodology because parents are frequently nearby during the interview.

Keystroke records for the 12 respondents who turned off the screen at least once during the interview were reviewed. This review proved quite interesting. In most cases, respondents turn the screen off for only a short time. In fact, the most common behavior was to turn the screen off and then back on within the space of a single question. One possibility is that the respondent turned off the screen while another person passed through the room where the interview was taking place. The ability to turn the screen off may have allowed respondents to protect the privacy of the answer they had entered for that question. Of course it is also possible that the respondent simply pressed the wrong key (for example, pressing the F7 key which turns the screen off rather than the 7 key to indicate using a drug 7 days in the past 30 days). This does not appear to be the case, however. When respondents pressed the F7 key prior to entering their answer, the answer entered was not a 7. And, in most cases, the screen was turned off after the respondent had answered the question. There did not seem to be an obvious pattern to the questions for which respondents chose to turn off the screen.

None of the respondents completed the entire interview with the screen off which could have been an indication of illiteracy. However, although it occurred infrequently, there were respondents who turned the screen off and left it off for more than just one or two screens. The largest number of screens for which a respondent chose to leave the screen off was 24. During the time the screen was off, this respondent appeared to provide consistent answers with no use of either the "Don't know" or "Refusal" keys. The interviewer made no special comments in the debriefing questions that could help explain the respondent's choice to answer questions without viewing the screen.

These results on the use of the "Screen Off" function are limited by the small number of cases and the qualitative nature of the analyses. However, these limited data suggest that the ability to turn the screen off may be preferred by some respondents. If the use of this function can further protect the privacy of the respondents' answers without adversely affecting data quality, as these results suggest, then honest reporting of sensitive behaviors may be further improved.

5. Timing Data

The time data that are automatically created by the Audio-CASI instrument can be used to pose a number of questions of the data from this CAI field test in order to examine how respondents may be differentially interacting with the Audio-CASI instrument. In the keystroke files we can identify which respondents used audio and video for the self-administered items, versus audio only. The time stamp data will reveal whether respondents are listening to the full question and all the answers categories for each question before providing their response. We expect that it is more likely that respondent use the audio to supplement their reading of the questions on the screen, and that they are unlikely to wait until the question is completely read through before answering. This has implications for questionnaire design. For example, those who use audio only (whether

for privacy or literacy difficulties) may only choose to hear certain parts of the question before formulating their answers.

In this section we explore the extent to which respondents are listening to all questions or whether they are answering before the question is fully read in Audio-CASI. To do so, we developed a "gold standard" measure of each question in both versions of the CAI NHSDA instrument. The gold standard is simply an independent measure of the time taken for the recorded voice to read the question through fully.

First, we examine the ratio of time taken for each question relative to the gold standard. A ratio of one means, on average, that respondents listened to the full question, no more and no less. A ratio less than one means that respondents are answering the question before the end of the question has been reached on the audio. A ratio greater than one means that respondents are listening to some or all of the response categories in addition to the question, or are taking time to consider their response.

Given that respondents in the SKIP version could be routed out of a section by giving a negative response to the first question, we focus only on the first question in each section, asked of almost all respondents. This avoids selection problems in comparing times on later questions. In addition, a varying number of outliers (defined as more than two standard deviations from the mean) were omitted in calculating the means.

The ratios of the mean times for the first question of each section relative to the gold time are presented in Table 3. The first thing to note is that there are few differences between the two versions of the instrument. There is a slight tendency for the SKIP version to have lower times than the MI, and we would have expected respondent fatigue or disinterest to have the opposite effect.

From this table, we do not see a steady decline in the ratios (as we might have expected) as respondents become more familiar with the questions and response process. In fact, it appears that the ratios may be determined more by the characteristics of the questions than by respondent familiarity or impatience. Thus, we see that the first question in the tobacco section has ratios greater than one for both the MI and SKIP versions. This suggest that, on average, respondents are listening to the full question, and apparently most of the response categories on Audio-CASI. Similarly, for the initial marijuana and crack cocaine sections, respondents are on average listening to at least the full question. However, for alcohol, cocaine and hallucinogens, respondents appear, on average, to be entering a response before the entire question is heard.

This variation can be further seen when we examine the individual items within the first section, tobacco. These ratios are presented in Table 4. Even within the first section, the ratios vary greatly, with some questions taking twice as much time on average than the gold time for the question stem (e.g., CG03, CG04) and others taking less time than the gold time (e.g., CG09, CG10). One suspicion is that the longer the answer categories relative to the question stem, the bigger the difference between good and poor readers may be in getting through the question. Also it may depend on usage, with those who have not smoked cigarettes more quickly able to find the category that fits their experience.

The data in Tables 3 and 4 suggest that we need a higher level of aggregation to explore subgroup differences in the average amount of the question listened to. To do so ,we created an average of the ratios of all questions answered within each section. This was done as follows:

Let i = person, k = question
Then $T(i,k)$ = time person i took to answer question k
$G(i,k)$ = gold standard time if $T(i,k) > 0$, otherwise
$G(i,k)=0$
If a $T(i,k)$ was an outlier as defined above, $T(i,k)$ and $G(i,k)$ were set to 0

Then the average ratios per section were calculated as follows:

$$\frac{\sum_i \sum_j T(i,k)}{\sum_i \sum_j G(i,k)}$$

In other words, it is simply the average of all ratios within a section, for those questions which were answered. The results of this summary approach are presented in Table 5. Looking first at the total population, we find more evidence of a decline in the time taken to listen to the question over successive sections. Thus, the average ratio for the 12 tobacco items is 1.14, while the average ratio for the 8 hallucinogen items is 0.40. This decline is monotonic over sections, suggesting that over the course of the Audio-CASI instrument, respondents may be listening to less and less of the question text.

However, the large variation across individual items suggests that there is some combination of question type and learning curve effects at work. Some questions cannot be easily answered before seeing or hearing the answer categories, while others may be immediately obvious to the respondent. On the other hand, the decline in portion of the question listened to may be a concern if important qualifying phrases appear at the end of a question or if important elements appear in the

response categories rather than the question stem.

6. Conclusions

Although these results are based on a small sample, they provide interesting insights into how respondents worked with the Audio-CASI system used in this field test. A second field test with a significantly larger sample size is currently underway and more detailed keystroke file analyses will be completed with this data. However, the data presented here provide initial support for the belief that respondents in this field test did not experience an inordinate amount of difficulty working with the automated NHSDA interview.

Acknowledgments

The authors are grateful to Michael Witt for his assistance in the preparation and analysis of the keystroke files.

References

Couper, M.P., S.A. Sadosky, and S.E. Hansen (1994). "Measuring Interviewer Performance Using CAPI," *Proceedings of the American Statistical Association Section on Survey Methods*, pp. 845-850.

Turner, C.F., J.T. Lessler, and J.W. Devore (1992). Effects of Mode of Administration and Wording on Reporting of Drug Use. In, Turner, C.F., J.T. Lessler, and J.C. Gfroerer (eds.), Survey Measurement of Drug Use: Methodological Studies. pp. 177-220.

Table 1: Percent of Respondents Who Used Special Keys

	Response Options	Audio/Screen Controls	Editing Keys	Escape Key	Sample Size
Mirror Image	20.7%	24.4%	74.1%	4.4%	135
Skip	18.9	23.0	61.5	4.1	122
Total	19.8	23.7	68.1	4.3	257

Table 2: Average Number of Times Special Keys Were Used

	Response Options		Audio/Screen Controls		Editing Keys		Escape Key	
	Avg.	% of Keystrokes	Avg.	% of Keystrokes	Avg.	% of Keystrokes	Avg.	% of Keystrokes
Mirror Image	5.8	1.0	9.1	1.5	7.5	1.3	1.7	0.3
Skip	3.5	1.0	9.6	2.6	6.1	1.7	1.4	0.4
Total	4.8	1.0	9.3	1.9	6.9	1.4	1.6	0.3

Table 3: Ratios of Mean Time to Gold Time for Total Sample, First Question for Each Section

		Ratio of mean time to gold time			Sample size		
Section	Question	Mirror	Skip	Total	Mirror	Skip	Total
Tobacco	CG-1	1.65	1.55	1.60	123	113	236
Alcohol	AL01	0.86	0.98	0.92	122	120	242
Marijuana	MJ01	1.30	0.98	1.15	131	117	248
Cocaine	CC01	0.61	0.77	0.69	132	112	244
Crack	CK01	1.39	1.06	1.34	127	25	152
Heroin	HE01	1.00	0.92	0.96	128	104	232
Hallucinogens	LS01	0.40	0.33	0.37	127	121	248

Table 4: Ratios of Mean Time to Gold Time for Total Sample, All Questions in Tobacco Section

	Ratio of mean time to gold time			Sample size		
Question	Mirror	Skip	Total	Mirror	Skip	Total
CG01	1.65	1.55	1.60	123	113	236
CG02	1.23	1.19	1.22	128	76	204
CG03	1.94	1.76	1.87	121	71	192
CG04	2.18	2.09	2.15	100	60	160
CG05	0.99	0.85	0.96	130	38	168
CG06	1.23	1.08	1.19	119	39	158
CG07	1.37	1.24	1.34	121	39	160
CG08	1.50	1.34	1.45	125	57	182
CG09	0.74	0.87	0.76	132	23	155
CG10	0.67	0.89	0.70	131	24	155
CG11	0.93	1.08	1.00	129	111	240
CG12	1.66	1.62	1.62	118	11	129

Table 5: Average Ratios Within Section of Mean Time for Total Population

	Mirror	Skip	Total
Tobacco	1.11	1.23	1.14
Alcohol	0.81	1.08	0.89
Marijuana	0.71	1.04	0.78
Cocaine	0.60	0.91	0.65
Crack	0.57	0.48	0.56
Heroin	0.51	0.98	0.55
Hallucinogens	0.42	0.36	0.40