Survey Methodology, June 1991
Vol. 17, No. 1, pp. 15–26
Statistics Canada

Assessing Measurement Errors in a Touchtone Recognition Survey

POLLY A. PHIPPS and ALAN R. TUPEK[1]

ABSTRACT

Electronic data collection utilizing touchtone recognition is in place for a monthly establishment survey at the Bureau of Labor Statistics. The Touchtone Data Entry (TDE) system features digitized phrases requesting respondents to answer questions using the numeric keypad of a touchtone telephone. TDE has substantial implications for lowering survey costs; many labor intensive activities are eliminated. However, little is known about measurement errors associated with this mode of data collection. This study assesses TDE mode error using three sources of data, which allow for analyses of errors associated with selected aspects of the human-machine interface. In addition, instrument design issues associated with mode error are addressed. We conclude by extending the implications of our findings to other surveys.

KEY WORDS: Mode of data collection; Human-machine interface; Computer-assisted self interviewing.

1. INTRODUCTION

The U.S. Bureau of Labor Statistics (BLS) issues monthly employment estimates for the United States from a survey of 350,000 business establishments. This survey, the Current Employment Statistics (CES) survey, provides one of the earliest monthly measures of U.S. economic health. However, the preliminary estimates from the survey are released with data from only about one-half of the business establishments in the survey. Revised estimates are produced two months after the initial press release. The low response rate for the initial press release can result in large revisions to the estimates. The BLS began investigating the use of automated collection techniques in 1983 to increase the timeliness of response and reduce the potential for large revisions.

The CES survey has traditionally been collected by mail through state employment security agencies. Research tests conducted between 1984 and 1986, involving the replacement of mail collection with computer-assisted telephone interviewing (CATI), have shown CATI to be an effective means for improving the timeliness of response (Werking, Tupek, Ponikowski and Rosen 1986). Average response rates under CATI collection have been between 85 and 90 percent for preliminary estimates, compared to 45 to 50 percent with mail collection. While CATI collection has been effective in improving the timeliness of response, the cost of full CATI collection in the CES survey cannot be absorbed within the survey's current budget. Research has been conducted since 1986 on touchtone data entry (TDE) to develop an alternative collection method with the performance gains of CATI, but with a lower unit cost (Ponikowski, Copeland and Meily 1989). Further discussion on the use of CATI and TDE collection for the CES survey is provided in the paper by Werking and Clayton, this issue.

Recent tests of the TDE system provide data on the timeliness of response, the cost of collection, and edit failure rates. BLS tests show that TDE collection with CATI back-up is as timely and effective as CATI collection (Werking, Tupek and Clayton 1988). In addition,

[1] Polly A. Phipps, Office of Employment and Unemployment Statistics, Bureau of Labour Statistics, Room 2821, 441 G Street N.W., Washington, D.C. 20212; Alan R. Tupek, Office of Employment and Unemployment Statistics, Bureau of Labor Statistics, Room 2919, 441 G Street N.W., Washington, D.C. 20212.

TDE lessens survey costs considerably. Current cost estimates indicate that the monthly unit costs for TDE are approximately 30 percent less expensive than mail collection, while the monthly unit costs for CATI collection are 20 percent more expensive than mail (Clayton and Harrell 1990). Collection by TDE is now expanding in the CES, beginning with the establishments with the greatest employment. Research continues in determining ways to improve the touchtone system, even though respondents' acceptance of touchtone collection is highly favorable.

The purpose of the current research is to identify respondents' problems using the touchtone system and to measure errors associated with the mode. Results indicate where improvements are needed to reduce respondent problems as well as errors. Section 2 provides background information on TDE and the operations of the BLS TDE system. In the third section we address the potential errors due to the TDE mode of data collection, which requires a human-machine interface. In section 4 we describe the three sources of data collected for the study. A record-check survey, machine-recorded data and a respondent debriefing survey are used in the analysis of problems and errors. Section 5 includes the methods, analyses and findings from each of the data sources. Finally, in section 6 we provide an overall assessment of measurement errors due to this mode of collection, suggestions for improving the system, and implications for other surveys.

2. TOUCHTONE DATA ENTRY

The primary reason to consider collection by touchtone self-response is to reduce the cost of collecting data by CATI, while maintaining the timeliness and quality of CATI collection. The CES survey seemed to be a good candidate for touchtone collection, since only five or six numeric data items are collected each month. The data items include: all employees, women workers, production workers, production-worker payroll, production-worker hours, and for some industries, over-time hours or employee commissions. Establishments are asked to report totals for each data item for the pay period which includes the 12th of the month. TDE respondents are also asked to report their establishment identification number and the month for which they are providing data.

The features of the BLS touchtone system include the ability to:

- detect legitimate establishment respondents based on a match to a file of establishment numbers;
- vary the set of questions depending on the industry of the establishment;
- read back all responses for respondent confirmation using a computer simulated (digitized) voice (respondents are requested to enter "1" to confirm their answer or "0" to reenter their answer);
- wait two seconds for respondents to begin entering their answer, and wait two seconds between digits before interpreting data entry as complete (data entry is also assumed complete if the entire field length is filled - example: the data item "month" has a two-digit field);
- repeat each data item question up to three times (for identification number, month and all employees) or request the respondent to confirm that they have no answer for the question (for all other data items), if a respondent does not confirm their answer or if no answer is provided in the two seconds after the question is read;
- store the date, start and end time of each call, and all data items (Werking *et al.* 1988).

Respondents are mailed instructions on touchtone data entry. The instructions give direction on how to use the system and examples of the computer and respondent interaction, such as:

Computer:	**Your Response:**
Enter all employees	For 25 employees, press 2 and 5
You entered 2, 5	Press 1 to confirm, 0 to reenter.

The instructions also include the optional use of the "#" sign to indicate completion of data entry for an item. The use of the "#" sign reduces the time of the interview. In addition, before reporting their data respondents can call in to try out the system using a special test identification number.

Touchtone respondents are contacted by a telephone interviewer during their first month on the system to determine any problems they may have and to provide guidance, if necessary. Respondents receive a postcard reminder each month and a prompt call if they do not self-report by a specified date. The prompt call asks the respondent to telephone into the touchtone system as soon as possible. Data are not usually collected during the prompt call.

3. MEASUREMENT ERROR IN A HUMAN-MACHINE INTERFACE

Respondent use of a system such as TDE to answer survey questions has little precedent. However, touchtone recognition is widely used in such procedures as electronic banking and customer-controlled telephone services. While these services may save time and expense, they have a potential to alienate users. Problems and errors can originate with the system, task, or respondent. System problems primarily generate nonresponse error, while measurement error is related to the task and respondent performance.

While not directly related to surveys, the human-factors literature suggests several interrelated factors that may contribute to performance errors in a human-machine interface. First, respondents may not be familiar or comfortable with the technology. Waterworth (1984) suggests that the language of the human-machine interface is different than human communication as actions are performed in an order reflecting computer program logic. Since the ability to think in a way that parallels the logic is not a minor exercise, those with limited experience may have difficulty understanding the task and using the system. Second, synthetic speech is more difficult to understand than natural speech and places greater processing demands on working memory (Schwab, Nusbaum and Pisoni 1985). Thus, comprehension and memory problems associated with the mode may cause errors.

Synthetic speech includes both digitized speech, where a human voice is sampled, digitally encoded, and stored, and rule-based synthesized speech, generated using text as input (Marics and Williges 1988). The TDE system utilizes digitized speech, which is less difficult to understand than rule-based synthetic speech. However, comprehension problems occur with digitizing, as it introduces distortion into original speech (Cox and Coope 1981). Research shows that the understanding of synthetic speech may improve with training. In an experiment on perception of synthetic speech, Schwab and colleagues (1985) found that training with synthetic speech increases perception performance. Thus, comprehension may improve with exposure to and experience with the system. Another factor that may affect comprehension is the pace of the system. Marics and Williges (1988) found that the rate of the synthesized speech significantly affects speech intelligibility, as measured by transcription errors and response latency. However, subjects who received contextual information prior to listening to the speech had fewer transcription errors.

Thus, potential errors in the human-machine interface can occur from lack of experience with the technology and task, and from comprehension and memory problems associated with voice clarity and pace. Yet these problems are surmountable, as the evidence indicates that experience and training can improve performance.

4. DATA

There were several objectives we considered in measuring TDE problems and mode error, and determining what data to use or collect. First, it was necessary to identify if and where problems were occurring. Second, we felt respondents should identify and interpret problems, but we also wanted measures independent of respondent assessment. Third, we needed to address problems and errors associated with the task and comprehension, including the possible improvement of respondent performance over time.

We decided to assess TDE problems and mode error using three different data sources, which have in common approximately 465 Pennsylvania business establishments. These establishments reported their monthly survey data by TDE to the Automated Collection Techniques (ACT) Laboratory at the BLS national office in Washington, D.C. A small number of the establishments began reporting by TDE to the ACT Lab in April, 1989. Others were added monthly through November of 1989. Most of the establishments continued reporting to the ACT Lab through April of 1990. The majority of these establishments moved from mail to TDE reporting.

The first source of data has two components. One is the TDE data recorded by machine from April to December, 1989. The other component is the same data recorded by establishments on a survey form. All respondents receive a yearly survey form on which they are requested to record their data for each month. Mail respondents fill in the form each month and mail it to the state employment security agency. The agency records the data, then returns the form by mail for next month's collection. CATI and TDE respondents are sent the survey form, but they do not return it. However, we sent a request to the TDE respondents to return their 1989 survey form, and obtained a 96 percent return rate. We then compared the TDE and form data, identifying discrepancies between the two. The TDE and survey-form data includes 1,930 observations across a nine-month period. Since establishments were phased into TDE slowly, the number of observations per establishment varies. The data cover approximately 75 establishments for 6-9 months, 200 establishments for 4-6 months and 190 establishments for 2-3 months. We refer to these data as the record-check data.

The second source of data includes machine-recorded information on respondent performance during the TDE telephone call. The TDE instrument was reprogrammed in January 1990 to automatically count and record the number of times a question was repeated due to nonresponse (question repeat), the number of times a respondent reentered data (data reentry) for each question, and the number of times an establishment called and hung up before entering data. Unfortunately, only the questions asking for the month and all employees total could be explicitly separated into question repeat and data reentry. For the data items including women and production workers, payroll and hours, we had to combine repeats and reentrys, due to the structure of the original computer program. We refer to these data as the machine-recorded data.

The third source of data is a telephone debriefing survey, conducted from January to April of 1990 with the Pennsylvania establishments on their experiences with the TDE system. Approximately 411 business establishment respondents completed the interview, an 88 percent response rate. The questions covered such topics as voice quality, pace of interview, task problems, use of systems features, adequacy of instructional materials, and a system rating.

5. RESULTS

5.1 Record-Check Data

When we requested TDE establishments to return their survey forms, our first question was: how many respondents really used the forms? We speculated that one source of mode error was respondents who did not complete the form for use when entering their TDE data, which would increase demands on their memory. Those who did not complete the form might be more likely to enter and/or verify incorrect data. Thus, the request for the survey forms indicated that respondents were to return the form regardless of whether they completed it or not. However, of the 96 percent that returned their survey forms, only one establishment mailed in a blank form; all others sent in completed forms. While nonrespondents may work from memory, most of the respondents had completed their forms, giving us reason to believe memory problems due to lack of form use were not a major source of errors.

When comparing the data received by TDE with that on survey forms, we identified and coded discrepancies. The data on the survey form are those we would have received and used if the respondents were reporting by mail. The results are shown in Table 1. The first type of discrepancy occurred when the TDE data indicated there was no response for a data item, but there was a response on the establishment's survey form. This item nonresponse accounted for the greatest number of discrepancies, 82 out of 177, and was quite evenly spread across the applicable data items.

There was a pattern to the item nonresponse by month and establishment. The item nonresponse rate was 40 percent higher in the first month an establishment reported by TDE. In addition, some establishments had more difficulty than others, indicated by two or more nonresponses. Nearly half of all item nonresponse occurred in 18 establishments at or close to the time they began responding by TDE. This indicated problems existed with first-time use of TDE that might decrease with experience. Since the problem was concentrated in a small group, we believe it reflected lack of familiarity with automated processes. The remaining item nonresponse had no identifiable patterns; our suspicion was that some establishments simply missed the item, possibly due to office distractions, and continued on with the next question.

Table 1

Record-Check Data - Number and Type of TDE Discrepancies

TDE item nonresponse	82
1-2 few/too many digits	18
Slipped on keypad	17
Dis/confirm - "1", "0" error	14
Form corrected, not TDE	12
No apparent error reason	26
Other reasons	8
Total	177

The second type of discrepancy was entering extra digits or, in a few cases, entering too few digits, which accounted for 18 of the 177 errors. This was specifically a problem associated with entering the payroll data item, where four respondents tried to enter cents instead of rounding to the nearest dollar. Several of the same respondents appeared to enter a half hour, 50, for production-worker hours rather than rounding. In the third type of discrepancy, the TDE numbers were nearly the same as those on the form, but one number off. The number entered incorrectly indicated a potential task problem in that the respondent may have had their fingers slide over on the keypad to the number directly on the side or below the correct digit. This accounted for 17 discrepancies. The fourth type of discrepancy occurred primarily for the all employee data item. There were eight establishments who had a "1" entered for this item in the TDE data, but had a larger employment number on their survey form. We speculate that respondents entered "1" twice when confirming the previous question on month.

Finally, there were a few respondents who had corrected data on their survey form, but not on TDE. There were other discrepancies which we could not explain. In addition, several respondents transposed their numbers or were off one category, accounting for the "other" reasons. For most of the errors, it was difficult to specifically ascertain if they were caused at the time of data entry, or not clearly comprehending the question or numbers being read back for verification. We suspected the former, but only for the second discrepancy, adding too many digits, could we really rule out comprehension problems.

The error rates for the survey items, ranging from 1.2 to 2.5 percent, are shown in Table 2. The all-employee, women-worker and production-worker questions have a lower percentage of errors than payroll and hours. This is not surprising since payroll and hours worked are usually four to six digits, compared to two to three digits for the other items. Thus, longer strings of numbers cause more difficulty. This may be related to difficulties entering the data, lack of respondent motivation in correction, or problems remembering longer strings of numbers during validation.

Table 3 shows the potential effect of the discrepancies on the CES data items, calculated by taking the sum of the difference between the values in the TDE system and the form, then dividing by the sum of the values on the form. The CES Survey uses a link-relative estimator for published estimates. The estimates in Table 3 do not take into consideration this estimator. However, the estimates in Table 3 provide an indirect measure of TDE mode error on survey estimates. None of the error is significantly different from zero at the five percent level. However, the potential for mode error appears to be more serious for production workers, payroll and production-workers hours. In this study, the number of production workers are overestimated by 7.3 percent, payroll by 7.3 percent, and hours by 4.4 percent.

Nearly all the discrepancies would have failed the edit parameters used in the CES survey and been corrected. The resultant effect of the discrepancies after edit corrections is zero for

Table 2

Record-Check Data - Number of Discrepancies and Percent Error by Data Item

	All Employees	Women Workers	Production Workers	Payroll	Production Hours	Total
Discrepancies	23	29	28	48	49	177
% Error	1.2	1.5	1.5	2.5	2.5	1.8
(SE)	(.2)	(.3)	(.3)	(.4)	(.4)	(.3)

(N = 1,930 for each item).

Table 3
Record-Check Data - TDE Mode Error for Data Items Before and After Edit Corrections

	All Employees	Women Workers	Production Workers	Payroll	Production Hours
% Mode error, before edit corrections	0.0	0.5	7.3	7.3	4.4
(SE)	(.4)	(.3)	(5.2)	(3.8)	(3.7)
% Mode error, after edit corrections	0.0	0.0	0.0	0.0	0.0
(SE)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)

(N = 1,886–1,930 for each item).

all data items, as shown in Table 3. Examples of the large discrepancies include a respondent who incorrectly entered payroll for the number of production-workers, increasing production workers by over ten thousand, and two respondents who incorrectly entered the number of production-worker hours for number of production workers, raising the latter by several thousand. Payroll and hours have similar gross discrepancies, including the respondents who put in cents, instead of rounding for payroll.

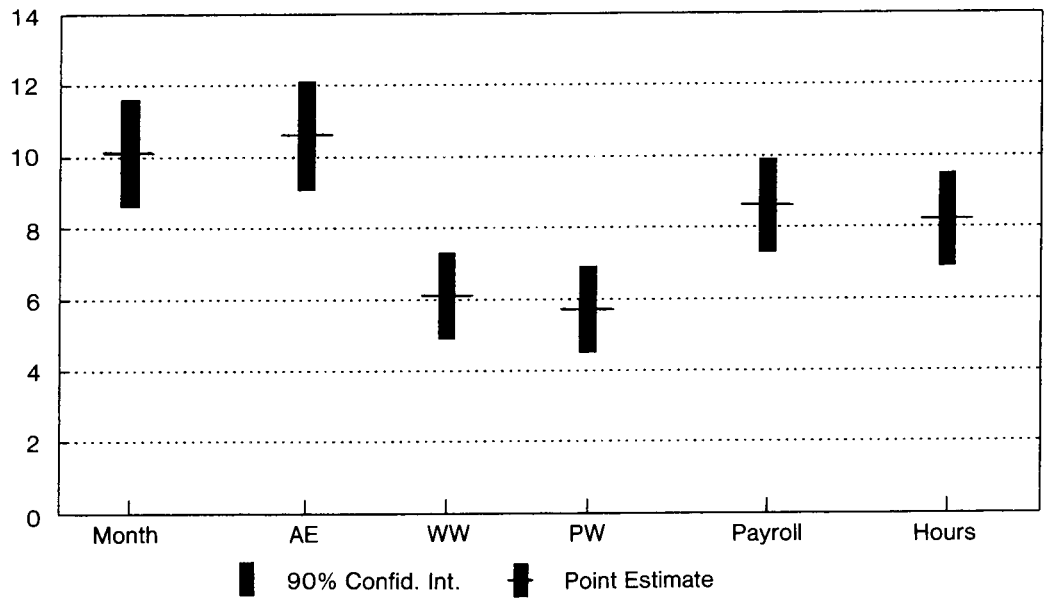
5.2 Machine-Recorded Data

The touchtone system provides a tool for assessing difficulties which respondents have with this mode of collection. TDE can record the number of times respondents reenter their data, and how often the question is read back a second or third time before they respond. Information can also be kept on those respondents who hang-up before entering data.

The machine-recorded data were collected for a total sample of 1,203 observations over a three-month period in 1990. There were approximately 474 unique respondents, many of whom provided data for two or three months. There were few differences in the machine-recorded data by month, so all data are presented for the three months combined.

Figure 1 provides, for each data item, the percent of calls for which the question was stated to the respondent more than once. The question could be stated a second time if the respondent does not answer in two seconds (repeat), or if the respondent fails to confirm his or her answer by entering "1," after it is read back (reenter). The figure indicates that the first two questions on the month and all employees, and the payroll and hours questions have higher rates of repeating and reentries by respondents than other data items. The higher rates for the first two questions - each over ten percent - may be due to respondents needing a few questions to orient themselves to the system. The payroll and hours questions generally have the greatest number of digits, so we suspect that data entry errors are more likely to occur, causing the question to be reread and the answer to be reentered.

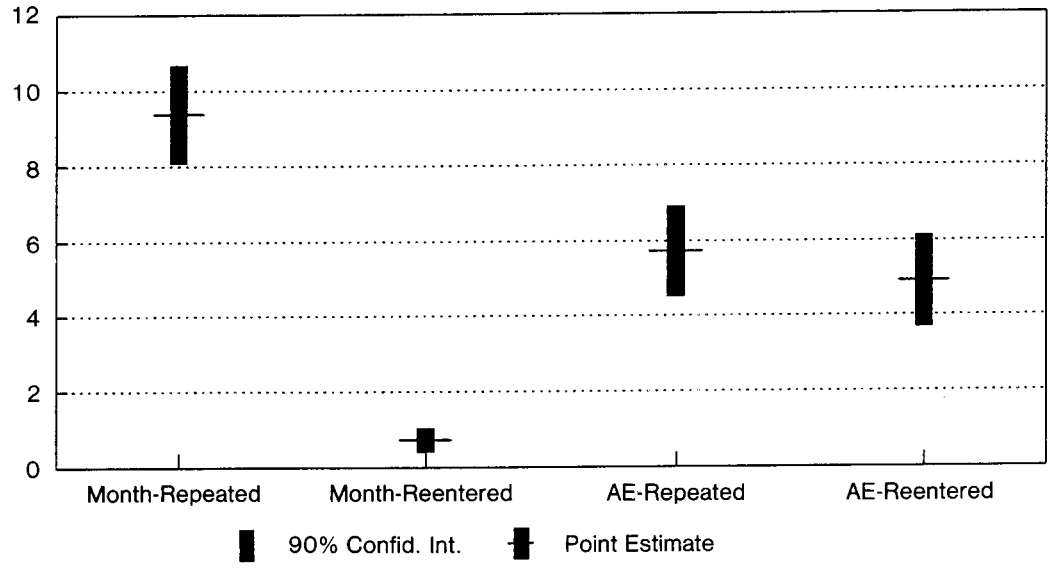
Figure 2 provides data for the first two questions on month and employment for repeated questions after no answer (repeat) and after lack of confirmation (reenter), separately. It was not possible to acquire data separately for the other data items. The CES touchtone system requires respondents to enter at least their report identification number, the month and employment. The system will accept item nonresponse for the other data items. The mandatory entering of month and employment allowed the separation of repeated questions after no answer versus after a respondents lack of confirmation of the answer provided. Of respondents with problems



AE = All Employees
WW = Women Workers
PW = Production Workers

(N = 1,203 per item)

Figure 1. Machine Recorded Data - Percent of Questions Repeated/Reentered



AE = All Employees
Repeated = No Answer Within 2 Seconds
Reentered = Correction of Answer

(N = 1,203 per item)

Figure 2. Machine Recorded Data - Percent of Questions Repeated Versus Reentered for 1st Two Data Items

on the month question, almost all were due to repeats, that is, two seconds passed without a response. On the other hand, problems with the employment question were almost evenly split between repeats after no previous answers and reentries after lack of confirmation of the previous answer.

Only two percent of the calls received by TDE each month included just a report identification number. During these calls, respondents had simply hung up, or could have been cut off the system.

In addition, the TDE system records all calls received that include at least the report identification number, month and employment. Using the TDE component of the record-check data discussed earlier, we identified respondents with more than one call during a month and coded reasons for the call-backs. In all, about four percent of the respondents called the system more than once in a given month. Most of these respondents provided data items which were not supplied on the initial call (2%). An additional one percent provided corrections to some data items in addition to new data items. Many of these respondents appeared to have had problems with entering the data the first time. Another one percent of the respondents called back only to provide corrections to data items previously supplied or provided identical data. These calls were often several days later, possibly implying that new data had been obtained from their records. In the case of the identical data, respondents may have forgotten whether or not they had previously reported their data. The system currently accepts the data with the latest date and time, although analysts are provided a list of respondents with duplicate records for review, and if necessary, correction.

A common reason for callback seemed to be related to the "enter 1" to confirm after each data item is entered. Many of the respondents who corrected their data had "1" in the data field prior to the callback, and some other response afterwards. Callbacks were twice as common with first-time respondents on the touchtone system than for respondents who were "experienced" users.

A few respondents called in their data three times for a given month, and one respondent called in data four times. These respondents seemed to be having difficulties with the system, but finally reported all of their data correctly.

Overall, these data suggest that respondents are having some difficulties with the system (more than they admit to during the respondent debriefing interview). Some steps could be taken to help alleviate some of the problems. These include providing more time for respondents to answer, providing better instructions, and trying to improve the confirmation of data entry method. In addition, being able to go back to a data item might solve some of the problems.

5.3 Respondent Debriefing Survey

BLS interviewers conducted a telephone debriefing survey with TDE respondents during 1990. Given the human-factors literature discussed earlier, some of the questions focused on understanding and pace of the digitized voice. The results from the machine-recorded data showed a substantial number of repeats and retries of questions, thus, questions were developed to address that topic. In addition, respondents were asked to rate the TDE system and answer questions relating to systems design.

The results from the debriefing survey are presented in Table 4. Respondents expressed little difficulty in comprehending the digitized voice. About 97 percent said the voice was very understandable, and all respondents indicated that it was easy to understand the numbers as the voice read them back for confirmation. During the first two months of the survey, we asked respondents about the pace of the interview. Most of the respondents said the pace was about right (88%), although about ten percent felt it was too slow. Our suspicion throughout the

Table 4
Results of Debriefing Survey*

Voice understandable	97%
Easy to understand #'s read back	100%
Pace about right	88%
Never reentered numbers	60%
Never repeated questions	83%
Never had poor telephone connection	93%
Used speed enhancement feature	63%
Instructions adequate	98%
TDE experience very favorable	93%

* N = 411, except for the pace and question repeat items. Approximately 177 respondents were asked about pace and 209 were asked about the repeating of questions.

study was that voice comprehension was much less a problem than were difficulties carrying out the task. While it was difficult to separate out the two in the record-check data, the debriefing interviews lend support to our suspicion.

For task difficulties, 60 percent of respondents said they never had to reenter numbers, while most of the others indicated they had to reenter numbers sometimes. When asked the reasons for reentering numbers, a majority indicated they had accidently entered a wrong number. Others said they did not have enough time, were distracted, or entered their numbers too fast. In the latter several months of interviewing, respondents were asked about the repeating of questions (without reentering data). About 83 percent of the respondents said they never found it necessary to repeat questions. Of the 17 percent who repeated questions, the majority said they were distracted, while others said they did not have enough time.

Most respondents had little difficulty with telecommunications failure, as 93 percent said they never experienced a poor telephone connection when using TDE. Of the respondents who did get a poor connection, most said it happened only once. A large number of the respondents, 63 percent, used the pound sign, a feature of the system designed for speeding up the reporting of data.

Nearly all respondents said the instructions sent to them as they began TDE were adequate. Overall, respondents seemed satisfied with the TDE system - approximately 93 percent rated their experience using TDE as very favorable.

6. DISCUSSION

The data show few serious problems with the TDE mode of data collection. Record-check data indicate some item nonresponse error, which is associated with first-time users. Entering additional or incorrect digits appears to be the most serious problem affecting the data items. However, in a panel survey, longitudinal edit checks could reduce this error, as could logical edit checks in all surveys. In addition, the rounding of data needs to be addressed in respondent instructions. Both the record-check and machine-recorded data show that there are more difficulties with longer strings of numbers, probably in both entering data and verifying

incorrect data. The latter could indicate difficulty remembering longer number chains during verification, as comprehension of numbers appeared to be good, *i.e.* respondents said they easily understood numbers being read back for confirmation.

Record-check data show that establishments may have carried over their confirmation of the month into the all-employee question. In addition, the machine-recorded data indicate respondents often do not respond to the month question the first time it is asked. Since respondents appear to be using their survey forms as they enter data, it is likely that moving from the identification number at the top of the form, to the month and data items further down the form, they require extra time to locate themselves. This problem could be solved by placing all information that needs to be entered in one location on the survey form. This might reduce the number of question repeats for the "month" item and potentially lower costs by reducing the length of calls. Question repeats for other items might be reduced by giving respondents more time to respond, since they report they were distracted from the task. However, since most respondents feel the pace of the system is about right, and many are using the speed enhancement feature, adding more time could cause frustration. Probably little can be done to reduce the number of reentries, as respondents indicate they have entered a wrong number and need to correct it.

The data show that errors are reduced with experience. This indicates that a panel survey may be best for this mode of data collection. For surveys requiring numeric or yes/no responses, we believe touchtone also has great potential. The errors are not extremely serious, and respondents rate their experiences using TDE very favorably. TDE may be particularly attractive to business respondents, who can call at convenient times, rather than be interrupted by telephone calls requesting data. However, for some surveys, self initiation and the lack of human contact may be problems which would contribute to nonresponse error.

Although respondent acceptance of touchtone collection is very favorable, there are some steps which can be taken to make the system better. These include:

- giving respondents enough time to key enter their data, especially for the first few questions and those which have a long string of digits,
- investigating ways to improve the confirmation of data items, and
- providing longitudinal edit checks to detect reporting of dollars and cents and other gross errors. The edits could be built into the TDE system with appropriate questions/probes to respondents to correct or confirm their answers.

BLS has used touchtone collection with one other survey. This survey was a small sample follow-up of business establishments who had participated in a Survey of Employer Drug Assistance Programs in 1988. The follow-up survey in 1990 was intended to determine if any substantial changes had occurred in the percentage of establishments providing employer drug assistance programs over the past two years. These establishments were mailed a short survey questionnaire requesting numeric or yes/no answers and encouraged to report their data by touchtone telephone. At the end of the first several weeks of the survey, approximately 20 percent of the establishments had reported their data by touchtone, and an equal amount by mail. TDE was not used after nonresponse follow-up activities began – about two weeks after the initial mailout. The remaining data were collected by telephone (CATI).

We believe that other surveys with time dependent data can take advantage of the time and keypunch savings of touchtone data collection. The mode may communicate the importance of timeliness to the respondent. This paper indicates that measurement errors are controllable using touchtone collection.

Given the timeliness and lower costs of touchtone data collection, we expect it will be used more extensively in the future. We know of two current projects testing touchtone recognition in a survey setting. Statistics Canada is testing a touchtone data collection system for the Survey of Employment, Payroll and Hours. In addition, a touchtone system for a survey of AT & T customers is being developed at Bell Laboratories (Wendler 1990).

BLS is also experimenting with the use of voice recognition technology for data collection in the CES survey (see Winter and Clayton 1990). While touchtone telephones are increasingly available, we estimate that between 10 to 20 percent of our respondents do not have touchtone telephones. Once speaker-independent voice recognition technology reaches an acceptable level for the ten digits needed to report CES data, we expect users will prefer it over touchtone collection. Further work on measurement errors associated with voice recognition technology needs to be undertaken.

ACKNOWLEDGEMENTS

The authors thank Darrell Philpot and Henry Chiang for their assistance with this research.

REFERENCES

CLAYTON, R., and HARRELL, L.J., Jr. (1990). Developing a cost model for alternative data collection methods: Mail, CATI and TDE. Presented at the Annual Meeting of the American Statistical Association, Anaheim, California.

COX, A.C., and COOPE, M.B. (1981). Selecting a voice for a specified task: The example of telephone announcements. *Language and Speech*, 24, 233-243.

MARICS, M.A., and WILLIGES, B.H. (1988). The intelligibility of synthesized speech in data inquiry systems. *Human Factors*, 30, 719-732.

PONIKOWSKI, C.H., COPELAND, K.R., and MEILY, S.A. (1989). Applications for touch-tone recognition technology in establishment surveys. Presented at the American Statistical Association Winter Conference, San Diego, California.

SCHWAB, E.C., NUSBAUM, H.C., and PISONI, D.B. (1985). Some effects of training on the perception of synthetic speech. *Human Factors*, 27, 395-408.

WATERWORTH, J.A. (1984). Interaction with machines by voice: A telecommunications perspective. *Behaviour and Information Technology*, 3, 163-177.

WENDLER, E.R. (1990). Respondent-initiated computer-directed surveys. Presented at the Annual Conference of the American Association of Public Opinion Research, Lancaster, Pennsylvania.

WERKING, G.S., TUPEK, A.R., PONIKOWSKI, C., and ROSEN, R. (1986). A CATI feasibility study for a monthly establishment survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 639-643.

WERKING, G., TUPEK, A., and CLAYTON, R. (1988). CATI and touchtone self-response applications for establishment surveys. *Journal of Official Statistics*, 4, 349-362.

WINTER, D.L.S., and CLAYTON, R.L. (1990). Speech data entry: Results of the first test of voice recognition for data collection. Bureau of Labor Statistics, Washington, D.C.