

Developing an Integrated System for Mixed Mode Data Collection in a Large Monthly Establishment Survey

Richard J. Rosen, David O'Connell, Bureau of Labor Statistics
Richard Rosen, BLS, 2 Mass. Ave. NE, Suite 4860, Washington DC 20212

KEY WORDS: CASIC, self-response, data collection, Business survey.

Background: The Current Employment Statistics (CES) survey is a monthly panel survey of over 390,000 business establishments. The survey publishes key economic statistics including employment, average hourly earnings, and average weekly earnings for the nation, as well as by industry, state and area. The employment estimates are closely watched by businesses, financial markets and policy-makers as a leading economic indicator.

The CES is a time-critical survey. Each month, there are only ten to fifteen days to collect and process the data before the preliminary estimates are published. Historically, most CES establishments have reported data by mail. Response rates for mail average only 55% by the cut-off date for preliminary estimates.

In response to both internal and external pressures, BLS has revamped its data collection procedures. Internal pressures include the need to improve response rates, reduce revisions, and control costs. External pressures come primarily from survey respondents demanding easier less burdensome ways to report.

Conversion to Automated Collection: In an initiative to improve estimates, raise response rates and reduce program costs, over the past 10 years, the CES has developed and implemented a number of automated collection methods.

These include:

- CATI-Computer Assisted Telephone Interviewing
- TDE-Touchtone Data Entry
- EDI-Electronic Data Interchange
- FAX-Used both for data collection and messaging
- WWW-Internet World Wide Web
- VR-Voice Recognition

These collection methods have transformed CES from a paper-driven labor intensive environment, to a paperless computer-driven environment. It has also meant a transition from a single-mode collection survey to a mixed mode collection environment.

Collection Methods Defined: In this section we define the various collection methods.

- **CATI-Computer Assisted Telephone Interviewing:** Under CATI, an interviewer makes pre-scheduled calls to sample units to collect their

data. The figures are entered into an on-line system that edits the data. The data are stored in machine-readable form for estimation.

- **TDE-Touchtone Data Entry:** The TDE system allows respondents to use the number pad on their phone to report their data into a remote computer. The respondent receives a pre-recorded interview asking for each data item in term. The data are stored in machine-readable form for estimation.
- **EDI-Electronic Data Interchange:** EDI provides a means for the respondent to directly transfer their data from the central data base on their computer to a BLS computer. The transaction requires that the data be sent using a standard file format. The data are stored in machine-readable form for estimation.
- **FAX:** Uses of facsimile transmission both for data collection and messaging. BLS currently operates two FAX systems. One system FAXes out copies of the CES reporting form. Respondents fill in this form and FAX the data back. Presently, the data are key-entered. The second system provides messages to respondents. This system is used to send Advance Notice messages (in lieu of mailing postcards), and Nonresponse messages (in lieu of interviewer phone calls). Both system are highly automated and use a computer-generated image and broadcast FAX technology to send thousands of messages per hour.
- **WWW-Internet World Wide Web:** Use of the Web for data collection is one of the very newest technologies. Respondents can link directly to the CES Web site and report their data using an electronic form. We provide links to other BLS Web sites where respondents can obtain BLS statistical and other information.
- **VR-Voice Recognition:** CES has operated a pilot VR system since 1978. The system recognizes digit and key words (such as "yes" and "no). Sample units call the system and receive computer-generated prompts, and "speak" their data into the system.

Development and Integration Issues: One of the goals of mixed mode collection is to have the appropriate mix of collection methods which balances collection objectives (such as response rates), costs, and respondent preference. For example, many respondents might prefer to receive a CATI call each month to report their data. However, it would be cost-prohibitive

to collect large numbers of units on CATI for such a large sample size.

With a mixed-mode collection environment there are also a number of new challenges to face. These include:

- Developing and testing each collection method prior to implementation.
- Maintaining and enhancing an array of specialized data collection operations.
- Determining the most appropriate collection method for each respondent.
- Developing the management tools necessary to track each unit, spot potential problems, and take corrective action.
- Integration of systems.

In the initial stages, development and testing can consume considerable resources within the survey organization. Indeed the survey organization may not possess the required skills to develop the system, requiring the organization to either hire new staff with the needed skills or contract out the development.

After development and testing, each new system invariably must go through a “growth” period where users (both from within the organization and respondents) demand improvements to the system either to correct initial problems, make the system easier to use, or to expand its scope and functionality.

Once in production, managing the flow of units into the system become important. How do you determine which units report by which methods? As mentioned above, if this decision is left up to the respondent, this may not be cost-effective.

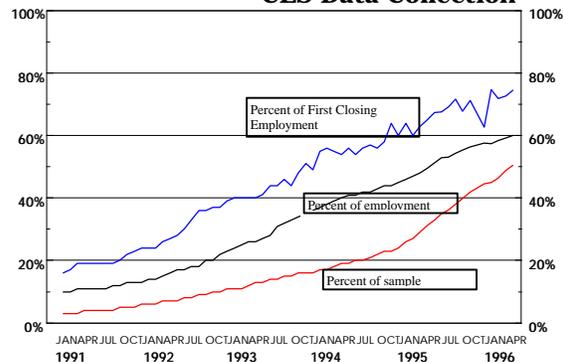
Managing the collection process then becomes paramount. For example, how do you keep track of which units are on mail and which are on TDE? Therefore, new management control systems must be developed for this and other purposes. With so many possible means to report data, the survey organization must develop mechanisms to control the assignment of collection method and the movement between methods. The survey organization must know which units are reporting by which methods. This is important for developing a total respondent contact system.

Finally, each individual system must be integrated into a total survey control system. While one may rightfully argue that this step should be performed up-front, in reality, until each system is operational, it is difficult to predict what the final product will look like, how many respondents will use the system, and what types of control feature will be needed. This makes design of a

System Control Module impractical until the systems “mature.”

Figure 1 shows the growth of automated collection methods in the CES over the past 10 years. It shows how the survey has moved away from the relatively slow labor-intensive mail environment to embrace TDE as the work-horse for data collection. Presently over one-half the CES sample report their data using their touchtone phone.

Figure 1. Contribution of CATI/TDE to CES Data Collection



One of the key issues is determining the most appropriate reporting method for each respondent. This decision must take into account such factors as:

- the technology available to the respondent (ie. do they have a touchtone phone, do they have Web access, etc.)
- the preference of the respondent
- cost considerations of the survey organization.

Presently, TDE is the most cost-effective reporting method for most CES sample units. The only requirement is that the sample unit has a touchtone phone. About 90% of businesses have a touchtone phone, therefore, our preference is to have as many units report via TDE as possible.

Profile of Collecton Methods: Figure 2 (show at the end of the paper) profiles the general uses, advantages and disadvantages of the various collection methods.

CATI: While CATI has the widest range of potential uses (from short to very complex questionnaires), it has one major disadvantage—cost. Because of the need to have an interviewer on the phone, the cost is considerably higher than the other reporting methods show. Thus, for large surveys, CATI may not be an option. CATI may need to be used for a segment of the sample population such as “critical” reporting units, or units that refuse to report by one of the more automated methods.

The other automated reporting methods shown have the potential to achieve high response rates (but not quite as high as CATI), with considerably less cost.

TDE: Because you are restricted to the number pad, TDE is limited to reporting numeric data or questions requiring categorical responses. You are also limited by the ability of the respondent to follow the branching sequence. In practice, this means that the survey designer should include only simple branching, otherwise the respondent may become confused and provide incorrect responses or terminate the interview in frustration.

Interview length may also be an issue. How long can a respondent be reasonably expected to stay on the phone with a computer? Can they continue to report accurately as the interview length grows? The average interview length to report data for a single unit for CES is about 1.8 minutes. Our research shows that respondents have little difficulty with a single report, and seem willing/able to report for several units at once. We generally restrict multi-unit reporting on TDE to units with 5 reports (about 8 minutes). Sample units who report more than 5 reports on TDE generally ask to be moved to another reporting method after a few months.

TDE's advantages primary come from its universality (90% of survey respondents have a touchtone phone), cost (a 2-minute computer interview costs about 15 cents), and the fact that the resulting data are easily stored in machine-readable form for direct input into editing/estimation systems.

Disadvantages include limited respondent feedback loops and limited editing capability. While collecting data from respondents via the computer is fairly straight-forward, providing information back to the respondent or answering questions become difficult. Respondents can be transferred to other automated response systems to provide additional information (i.e. such as wage rates in their industry), however, respondents may not enjoy being "bounced" from one system to another.

Editing is difficult because of the need to phrase questions that the respondent can understand (without a person/interviewer assisting) so they can provide appropriate responses. It would seem that editing would have to be limited to very basic logic/internal checks. The question thus becomes, is it cost-effective to built in such basic edits, given that fact that a full range of edits will need to be performed on the data at a later stage?

The BLS TDE system currently does not perform any edits. Our research shows that only about 3% of reports would fail the type of basic logic checks that we believe could be reasonably programmed into the system. Most of these failures can be cleaned-up by inspection of the data.

FAX: FAX has many of the same uses as TDE and offers many of the same advantages as well. Respondents can be FAXed simple survey/data collection forms to complete and FAX back to the survey organization. FAX is almost as universally available within the business community and appears to be growing. About 80% of firms with more than 10 employees have a FAX machine, not much different than the 90% that have a touchtone phone.

The big disadvantage of FAX is the difficulty in processing survey results. Despite advances in character recognition (CR) technology, reading handwritten/printed responses is only about 90% accurate. This means that numeric responses, such as those required for CES reporting, cannot be automatically processed through CR systems without significant human intervention. Results from the Census Bureau on CR on forms similar to the CES found recognition rates of about 90% for each character. For a typical CES form with 5 data items of 3 character each, this results in a total recognition rate per schedule of about 18%. Thus, a typical CES form may have several occurrences of non-recognition requiring manual review. For surveys with check boxes or other categorical responses, CR can provide very reliable recognition and can be more beneficial.

Despite the problems related to recognition, FAX should be considered a viable means for survey reporting. Even if the form must be completely key entered, collection costs will be lower than mail for short forms, and FAX offers many benefits related to timeliness. For example, a one page FAX can be sent to a respondent at a cost of only 7 cents compared to 32 cents for mail-out. There are also savings associated with printing and preparing mail-out, since a few key-strokes on a computer can run a program and generate thousands for FAX messages for transmittal. These savings more than make up for the additional cost of key entry or verification/review of CR edit failures.

EDI: Although EDI has many cost advantages, it is only appropriate for companies that provide large volumes of data. This volume is needed for the firm to justify the up-front cost associated with programming and testing the application. Companies that provide only a few reports would not likely be willing to invest

time to program such an interface. In addition, EDI is only applicable for reporting of numeric information and there is no opportunity for editing. Also, there is little avenue for respondent feed-back, since in most instances, there is simply a computer to computer transfer of information. For CES, we estimate that only about 5% of sample units will be willing/able to provide data via EDI. Still, for these firms, EDI has the benefit of substantially reducing their reporting burden.

Web: Use of the Web offers one of the most powerful data/survey collection tools for survey organizations. It provides the most comprehensive array of potential uses, from simple to complex surveys, as well as major cost benefits. Perhaps its most powerful benefit is in the area of customer interface. The Web combines an array of graphic interfaces not available with other collection vehicles. It has the potential to transform survey collection into an interactive experience. The Web can be used not only to collect information from respondents, but provide information back to respondents in a fast, efficient, and user-friendly manner. This should help solidify the reporting arrangement and reduce attrition, important considerations since most survey's (including CES) are voluntary.

Presently, the greatest drawback to Web collection is the limited access by most survey respondents. Our research shows that at the present time only about 15% of CES respondents have the required software and Internet access. However, Internet access is growing rapidly, and we can expect that over the next 5 years many more survey respondents will be able to report via the Web.

Another issues related to the Web is data security. Most survey organizations collect information under a pledge of confidentiality. Advances in Web security and encryption techniques are rapidly addressing these concerns and should make the Web as secure as other data reporting methods.

Non-response Issues: Non-response Prompting (NRP) is critical to the maintenance of high response rates. For CES, in a typical month, 35-40% of the sample do not report by our suggested "due date" and receive a friendly reminder prompt. At present, depending on the size of the firm and availability of FAX, this prompt message may be via CATI, FAX, or postcard.

As with data collection methods, there are a number of choices for conducting nonresponse follow-up. These include: CATI; FAX; E-mail; and Mail.

Each method has its advantages/disadvantages and cost tradeoffs. For CES, we have found that, of those units

that are prompted, approximately 60% report data prior to the deadline. The effectiveness of each prompting method is not uniform and must also be considered, especially if response rates are a primary concern.

CATI: CATI would appear to be the most effective method of performing nonresponse follow-up since it provides direct contact with the respondent. However, as with data collection via CATI, this comes at considerable expense. An experienced interviewer can complete 20-25 call per hour. At an hourly rate of \$20.00, the labor cost for NRP calls is \$0.80. Add to this the cost of the two minute phone call at \$.15 and the direct cost per call is \$0.95. The experience in CES is that 60-70% of the units prompted via CATI will report their data by the suggested deadline.

FAX: Since 1984 BLS has operated a broadcast FAX message system which has the capability to send thousands of FAX messages to respondents. The operator presses a few keys on a computer, and the FAX system sends a customized message to each sample unit. Thus, the only direct cost is the telephone charge for connecting. The one-page FAX which we currently send takes less than one minute. Our current phone charge is less than \$0.10 per minute. However, because there is no direct contact with the establishment or the contact person, we have found that FAX is somewhat less effective than CATI in eliciting a positive response from the sample unit. Our experience is that 50-60% of those prompted via FAX will report by the suggested deadline.

E-mail: E-mail provides yet another level of both speed and cost savings. Broadcast E-mail can be accomplished with a few key strokes and the messages are received in moments. As part of Web collection, respondents are sent advance notice messages via E-mail, and, if they have not reported by our deadline, they receive a nonresponse e-mail. E-mail messages are virtually free. This is because, a large survey organization will likely have a blanket fixed price contact for e-mail service. The cost of this service is spread throughout the survey organization, there is no fee per message, and the marginal cost of an additional message is negligible. Our experience for E-mail NRP is that its effectiveness is about the same as FAX, making this an excellent choice for NRP for units that have E-mail accounts.

Mail: Nonresponse prompting via mail is not effective in a time-critical survey. Since it takes several days for mail to reach the respondent, nonresponse activities must begin very early in the collection cycle. Thus a larger proportion of the sample will need to be prompted if NRP is performed by mail. Indeed, many

units sent an NRP message via mail will likely report prior to receipt of the mail reminder. For surveys with broader time spans or situations where a substitute survey form must be sent, mail is an option.

Cost Issues: Survey organizations are constantly striving to control/reduce costs. Except for CATI, the other automated collection methods offer varying degrees of cost saving over mail collection. Figure 3 reviews the major cost components for each collection method.

While CATI is the most expensive mode of collection (outside of personal visits), it does yield the highest response rate. If one recalculates the unit cost taking into account the response rate, CATI's cost penalty is greatly reduced. This "response rate adjusted" unit cost is also shown in Figure 3. Under this basis, the cost of mail increases by 100% (since mail yields only a 50% response rate); however, CATI costs increase by only 10% (since CATI yields a 90% response rate). TDE, FAX and Web continue to exhibit major cost advantages over mail and CATI. EDI's cost advantage is greatly reduced based on the response rate adjusted basis. This is because of the significantly lower response rate for EDI. However, the basic premise of EDI is to reduce respondent burden, and EDI may be the only reasonable means to collect these data from large units.

Converting a large share of the sample from Mail to TDE produces significant ongoing cost-savings. Most of the unit cost for Mail is for labor and postage, the prices of which continue to rise. With TDE, most of the unit cost is for telephone service and computer hardware, the prices of which continue to fall. In addition, data entry is eliminated under TDE. The two-way first class postage for the report form is replaced by one-way postage for a TDE postcard, or the even more cost effective FAX or E-mail. TDE also eliminates the monthly labor costs of opening, stuffing and sealing envelopes. The TDE FAX messages have nearly zero labor costs and also take advantage of the falling costs of telephone calls. (Clayton and Harrell, 1989). The cost for E-mail is virtually zero.

If one constructs a cost-benefit ratio for each method, it becomes clear that the marginal advantage of CATI NRP in increasing response rates is more than overshadowed by the high cost-benefit ratio. For CES, CATI nonresponse follow-up raises overall response rates by only 2-3 percentage points compared with FAX or E-mail. Thus, unless the survey organization determines that this increase is imperative, they should consider one of the more automated NRP options. Indeed it could be argued that the additional expense of

CATI nonresponse may be better spent on other survey activities.

Figure 3. Monthly Unit Cost of Data Collection-Adjusted for Response Rate Differences Ongoing Transmission and Data Entry Costs

Item	Mail	CATI	TDE with FAX NRP	TDE with Phone NRP	FAX	EDI	Web
Phone Charges	--	\$ 0.88	\$ 0.28	\$0.16	\$0.08	\$ 0.28	\$ 0.13
Postage	\$ 0.76	0.23	--	0.20	--	--	--
Labor	0.29	1.10	0.04	0.20	\$0.35	0.01	--
Total	\$1.05	\$2.21	\$0.32	\$0.56	\$0.43	\$0.29	\$0.13

Response Rate	50%	90%	75%	80%	75%	60%	75%
Adjusted Unit Cost	\$2.10	\$2.45	\$0.43	\$0.70	\$0.57	\$0.48	\$0.13

Summary: The technology available to the survey organization has greatly expanded the potential methods of contact (both initial contact and nonresponse follow-up) with respondents. For CES, conversion of Mail reporters to TDE has produced desirable long-term results including higher response rates, lower attrition, and reduced revisions. Over 250,000 units are currently on TDE. Web reporting offers the potential for further cost savings without sacrificing response rates. Indeed, Web reporting has other benefits in the area of customer satisfaction.

Survey organizations must constantly strive to both improve/control costs, maintain/improve response rates, maintain/improve data quality.

References:

Rosen, Richard J. (1991), "Improving Nonresponse Follow-up in a Monthly Establishment Survey", Proceedings of the Section on Survey Research Methods, American Statistical Association, 587-593.

Clayton, Richard and Harrell, Louis (1989), "Developing a Cost Model of Alternative Data Collection Methods: Mail, CATI and TDE," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 264-269.

Werking, George S., and Clayton, Richard L. (1995) "Automated Telephone Methods for Business Surveys" in *Business Survey Methods*. New York: Wiley-Interscience. pp. 317-338.

Figure 2. Profile of Collection Methods

	CATI	TDE	FAX	EDI	Web
Potential Uses	<p>Structured Interviews</p> <p>Simple survey to complex branching & editing</p> <p>Interview length is flexible (short or long)</p>	<p>Structured Interviews</p> <p>Simple survey with limited branching</p> <p>limited interview time</p>	<p>Structured Interviews</p> <p>Simple survey with limited branching</p> <p>limited survey/page length</p>	<p>Collection of large volumes of data</p>	<p>Structured interview or data capture</p> <p>Branching limited only by respondent's ability to comply</p> <p>Length of interview unknown</p>
Advantages	<p>Direct contact with respondent</p> <p>Computer allows for complex branching</p> <p>Data can be edited and verified</p> <p>Data stored in machine-readable form</p>	<p>Low cost</p> <p>90% of population can use the system</p> <p>Data stored in machine-readable form</p>	<p>Low cost</p> <p>80% of population have FAX</p>	<p>Low cost</p> <p>Data stored in machine-readable form</p>	<p>Low cost</p> <p>Easy customer interface</p> <p>Obtain customer feedback</p> <p>Provide specialized information to respondent</p> <p>Flexible</p> <p>Ability to edit data</p> <p>Data stored in machine-readable form</p>
Disadvantages	<p>More costly than other reporting methods</p>	<p>Restricted to numeric data or categorical responses</p> <p>Limited editing capability</p> <p>Little opportunity for respondent feedback</p>	<p>Data must be key entered or need to develop OCR system</p> <p>No editing capability</p>	<p>Considerable up-front work required by respondents</p> <p>No editing</p> <p>No opportunity for respondent feedback</p> <p>Not applicable for questionnaires, only data</p> <p>Only 5% of respondents willing and able to do EDI</p>	<p>Only 15% of respondents have required HW/SW for Web access</p>
Response Rate	90%	80%	75%	60%	80%