Chapter 11

CATI-CAPI Conceptual

Chair: M. Denice McCormick Myers, National Agricultural Statistics Service

R. Jamieson

David O'Connell

Mark Pierzchala

Chapter 11

Statistics Canada's Experience in Moving to CAI from Paper and Pencil

R. Jamieson, Statistics Canada

Abstract

In Survey Operations we have been using Computer-Assisted Interviewing (CAI) to improve quality and reduce cost for several years.

In the past 3 years we have converted 14 economic surveys from paper and pencil mode to CAI. By the end of January 90% of our surveys will be using this mode of collection for annual, quarterly and monthly data.

These interacting applications have been developed to handle both telephone and the mail components for each survey. The edits include the standard interfield, historical, tolerances, balancing, and consistency.

The presentation will include:

- description of the collection process
- description of the type of edits we are using for 3 applications, including the Consumer Price Index
- the impact that moving to CAI has on the operations
  - resources
  - interviewer training/manuals
- application development
  - problems
  - sharing code. ■

Chapter 11

A Feasibility Test of On-Line Editing for Touchtone Data Entry Collection

David O'Connell, Bureau of Labor Statistics

Abstract

The Current Employment Statistics (CES) Survey is a monthly panel survey of over 390,000 business establishments. Nearly 40 percent of the units in the CES sample report their payroll data by telephone to the survey's Touchtone Data Entry (TDE) system. This automated data collection system has made data available faster and has reduced program costs. Because TDE captures the data in machine-readable form that can be easily fed into an editing system, the CES is working to integrate an editing system into the data collection system. With such a system, a series of edit checks could be run on data as they are reported. In the event that any items fail edit, appropriate feedback could be provided to the respondent. Such a system could speed up the editing process by resolving many edit failures automatically and thus reduce the editing workload of the human analysts.

Two approaches are under consideration:

- The first approach is an on-line edit message. If a report fails one of the basic edit checks, the system would speak a short message to the respondent. This message would explain the edit failure and ask the respondent to check that they entered their data correctly. The system would then allow the respondent to re-enter their data items or to add enter an explanation code.

- The second approach is a "FAX-back" system. For any reports failing one or more checks, the system sends a FAX message to the respondent. This message explains the nature of the edit failure and asks the

Abstract (Cont'd)

respondent to check their data and either call in a corrected report or FAX back an explanation. Because it is printed, the FAX would be able to explain more sophisticated edit failures than would a short spoken message.

For either approach, there are, however, a number of conceptual and methodological questions which must be addressed. These include:

- What type of question scripting/branching should be used?
- How sophisticated can the edits be?
- How will respondents react to these types of questions?
- Are respondents sufficiently knowledgeable about the data to provide accurate responses?

The paper provides a full review of these issues as well as the results of the first feasibility test of the FAX-back system.

A Feasibility Test of On-Line Editing for Touchtone Data Entry Collection

David O'Connell, Bureau of Labor Statistics

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. Average response rates for mail are only 55 percent by the cut-off date for preliminary estimates. Because of the low response rate for mail, the preliminary estimates have been subject to considerable revision. Furthermore, collecting data by mail is labor-intensive and costly.

Conversion to Automated Collection

In an initiative to improve estimates, raise response rates and reduce program costs, the CES has developed and implemented a number of automated collection methods.

In 1984, CES began to develop *Computer Assisted Telephone Interviewing* (CATI), in which live interviewers call respondents and collect their data over the phone. CATI regularly achieves response rates in excess of 90 percent, but because it is relatively expensive, it is used by CES primarily as a transition mode. Larger reporters -- 50 or more employees -- spend six months on CATI and then are converted to automated self-response. Smaller reporters are converted with a single telephone call without collecting data.

Most reporters are converted to *Touchtone Data Entry* (TDE), a system CES started developing in 1986. Under TDE, the respondent initiates a phone call to a computer system and enters data directly using a touchtone telephone. Data are available to the survey in minutes instead of days, and in machine-readable format. The costs of postage, handling and key-entry are greatly reduced compared to mail, and self-response makes it much cheaper than CATI. Respondents have been very receptive to this collection mode, producing average response rates of 80 percent. Over the past decade, CES has converted 47 percent of the sample to TDE.

CES has also developed a number of other automated self-response data collection systems for reporters meeting special criteria. For respondents without touchtone service, CES developed a *Voice Recognition* (VR) system that recognizes spoken responses to the automated interview. For respondents who report for dozens of establishments each month, *Electronic Data Interchange* (EDI) enables them to transmit all of their reports at once quickly and easily, computer-to-computer. For respondents with Internet access, CES developed a secure site on the World Wide Web in 1996 that collects data and provides interesting and useful feedback to our respondents.

Features of Touchtone Data Entry

Over 170,000 units currently report by TDE, with over five thousand new units converted from Mail each month. Each TDE reporter receives a report form and instructions.

In the middle of each month, each reporter gets an "advance notice" message either by FAX or postcard, which shows suggested reporting dates and the toll-free number. When the establishment's payroll data are available, the respondent initiates a call to the computer and is guided through a brief interview. The respondent enters data by using the numeric keypad. Each response is repeated for verification. The average interview lasts two minutes, and the respondent may call at any time of the day. As the reporting deadline approaches, missing reporters are prompted with either a phone call or a FAX.

Editing Microdata

CES microdata are edited for validity, for adherence to tolerance ranges, and for month-to-month consistency. These edits are currently performed on the mainframe in batch mode. Any report that fails edit must be coded or corrected by the data analysts. Performing edits in this manner makes sense when reports are collected by mail and must be key entered by the survey's data entry staff. It makes less sense in the TDE environment, when the data enter the survey office in machine-readable form. Editing these reports on-line may achieve significant savings.

Edit reconciliation is a labor-intensive, time-consuming activity. Analysts review the editing output manually, contact the respondents when necessary, correct errors and code valid reports. There is often a considerable time lag between the time the respondent reported the data and the eventual editing follow-up contact. In addition, most "failed" data are found to be correct. Time constraints limit the amount of review time available before the initial estimates are published, and the latest data received are least likely to be edited.

TDE collection may enable the CES to address some of these editing challenges through automation. Automatic on-line editing would treat each data report as it is entered and notify the respondent immediately in the event that a data item fails edit. This notification could either be part of the interview that the respondent hears, or it could be a FAX message, generated during the interview and sent to the respondent within minutes.

One benefit of hearing an editing message as part of the TDE interview is that the respondent can make corrections or enter explanation codes at once during the same call. Possible drawbacks include the increased length of the interview and the possibility of confusing the respondent. In the event that the respondent needs to check other records or recalculate the data, it is unlikely that the editing question would be resolved in the same call.

The FAX message option would provide the respondent with a clear, printed version of the editing question. The data for the current month and previous month could be shown with the questionable items highlighted. Because some edit failures are complex, it may help to see such a message as opposed to hearing it. The message would ask the respondents to check their data sources and then either phone in a corrected report or send an explanation by FAX. Some potential drawbacks of the FAX message are that the FAX may not get to respondent in a timely manner, the confidentiality of the data may be impaired (within the respondent's company), and those explanations returned by FAX would require further attention.

The CES program has found that FAX is widely available. Eighty-five percent (85 percent) of CES respondents have FAX machines, and this medium is accepted for receiving important messages. Currently, CES uses FAX to send monthly messages to its Touchtone respondents for advance notice and non-response prompting. FAX is also used to send CES informational materials during solicitation and Mail-to-TDE conversion.

Feasibility Test

Before proceeding with developing an on-line editing system, certain methodological concerns must be addressed. The respondents' reaction to receiving on-line edit messages is the most important factor in determining whether this is a useful area for development. Therefore, an initial feasibility test was conducted to determine the following: respondents' willingness to receive this type of message; their ability to understand the message; their responsiveness to the message; their ability to provide corrections or explanations; and the timeliness of their responses.

The objective of the feasibility test was to determine TDE respondents' reaction to on-line edit messages. The approach selected was to use:

- edits in the CES CATI system to edit recently entered TDE data,
- a word processor to customize editing messages, and
- a FAX machine to send the messages to the respondents.

Therefore, the FAX editing option was the one that was tested.

The test sample was selected from the TDE units in our Atlanta Data Collection Center. During the test period in May 1995, 400 units reported data. These reports were edited with the CATI system, and approximately 80 failed at least one edit. Of these, 20 met the research criteria of having a valid FAX number, an edit failure that was not resolvable by human inspection, and a generally "good attitude" toward the survey (respondents who seldom needed prompting and had been receptive to past editing phone calls).

We experimented with three different treatments. In the first treatment, the analyst called five respondents to resolve the edit failure. They were informed them that we were sending a sample of a FAX message that we were considering in place of such calls, and we asked for their reactions. The feedback was positive. Respondents said that they understood the message and would view it as important. Respondents said they would respond to such a FAX within a day or two, and were very willing to FAX-back responses. One respondent expressed some concern about the fact that the BLS FAX message showed the company's payroll data, which could be viewed by other employees in her immediate office.

In the second treatment, six respondents were called only to notify them that we were sending them a FAX message about their data. Then the message was FAXed and we awaited responses. Of the six reporters, three FAXed explanations; one called us to say that she refused to check her data; and two did not respond.

In the third treatment, 9 reporters were sent the FAX message without pre-notification. Of these, seven FAXed back explanations; one phoned in a corrected report; and one did not respond.

Summary

Combining the results of the 2nd and 3rd treatments, 80 percent (12 of 15) responded to the message in some way: ten by FAX; one by phoning in corrected data; and one by telephone (a refusal). All responses came within three working days. Most FAXed explanations were sufficient for the analyst to code the data. Two required further explanation by phone. While not statistically significant, the results of this test suggest that some of our TDE respondents would be willing to receive FAX editing messages, would understand them, and would provide useful, timely responses.

Future Developments

The results of the first feasibility test are encouraging enough to consider further work in this area. One of the next logical steps is expanded testing, to obtain statistically significant results. Another is to truly automate the process by programming various edits into the TDE system and utilizing the system's existing FAX capability to create and send FAX messages automatically. The messages themselves also must be refined, so that they are clear and appear official. The data confidentiality issue must be addressed, perhaps with the use of a "consent" flag. Finally, we must consider ways to handle the responses the system will generate. Because of its potential to save considerable resources and improve the quality of data collected, on-line data editing may one day become an integral part of the TDE system. All of the lessons learned through these tests are also contributing to on-line editing screens and prompts in the Internet collection system. ■

Note: Any opinions expressed in this paper are those of the author and do not constitute policy of the Bureau of Labor Statistics.

Chapter 11

CAI and Interactive Editing in One System for a Survey in a Multi-mode Environment

Mark Pierzchala, Westat, Inc.

Abstract

The coordination of Computer-Assisted Interviewing (CAI) and interactive editing for one survey is discussed. Computer-Assisted Interviewing seeks to obtain clean data at the time of data collection. CAI controls the routing depending on the responses recorded during the interview, ensuring that interviewers ask appropriate questions. Other kinds of edits, for example consistency edits, can be placed into the interview and reconciled with the respondent. Interactive editing is a form-by-form review of all edit failures after data collection. It speeds the review by giving immediate feedback to data editors on their actions, allowing them to clean a form in one session. It eliminates the hand inspection of paper forms before data entry and the need for keypunch of corrections after the edit. Interactive editing may take place after CAI or after data are gathered by paper questionnaires. A third form of data editing, top-down editing (also known as macro-editing or statistical editing), is briefly discussed, including its effect on how interactive editing is done. This paper is based on work done in the Blaise system from Statistics Netherlands on the large and complex June Area Frame survey conducted by the National Agricultural Statistics Service (NASS).

In developing a dual-mode instrument for both interviewing and data editing, developers must keep in mind the differing needs and preferences of the interviewers and the data editors. In NASS, it is necessary to additionally adhere to processing conventions that were developed over the years. These conventions were developed separately for batch processing of paper-collected data and for Computer-Assisted Telephone Interviews (CATI) using different systems. Blaise effectively and efficiently integrates these two modes of survey data processing in most respects. In NASS' implementation, there is some additional design and programming work needed to mediate between the two modes of use, due to the previously adopted conventions. On the other hand, the use of one system for both modes eliminates much double programming that is done when two different systems are used and presents a large net savings in preparation effort. There are also some methodological gains.

CAI and Interactive Editing in One System for a Survey in a Multi-mode Environment

Mark Pierzchala, Westat, Inc.

The June Area Frame Survey

The June Area Frame survey is conducted annually in all of the 48 continental states in the United States. In a two week period in early June, 120,000 people are visited by 1,500 interviewers in almost every county in the United States. About 50,000 of these contacts result in full interviews. By the end of June, livestock inventory and grain stocks are released, with other commodity estimates following in July and August. Additionally, subsampling for down-stream surveys starts around June 20. To realize this statistical production schedule, a clean data set must be produced within a few days of the end of data collection.

Interviewers visit land areas called segments, with an aerial photograph in hand. Together with the respondent, they draw off land in the segment operated by that person on the aerial photo. This is called a tract. If the tract is agricultural, they conduct a full interview. They mark off fields in the tract on the photo, and determine what is in each field. As interviewing is done at the height of the planting season, the farm respondents are usually very busy. Interviews are conducted in crop fields (much dirt and dust), brilliant sunshine, blistering heat, perhaps in a downpour, or in the barn (more dust).

The June Area Frame dual mode interviewing and editing instrument is complex, encompassing five levels of organization (segment, tract, field, crop, crop utilization), with over 7,000 possible questions (usually about 100 are asked), 8,900 edits, and thousands of computations and routing instructions. There can be a maximum of 78 tracts per segment and up to 99 fields per tract though most tracts have fewer than 20. Within a field, there can be an indefinite number of crops, though it is extremely rare to have more than two. There are 44 versions of the June Area Frame survey -- one for each of NASS' state offices. Each state has its own agricultural commodities that are surveyed, edit limits, question wording, and units of measurement. One dual mode instrument is driven the 44 different ways, at runtime, by an external Blaise database of state-specific meta data. It takes minimally 800 pieces of information to specify a state's version, and up to 1,200 specification items for more complex states. Fortunately, NASS has organized this so that every state's questionnaire can be thought of as a variation on one master theme.

NASS conducts this survey on paper but started a small Computer-Assisted Personal Interviewing (CAPI) -- three interviewers -- and interactive editing project (1/2 the sample) in Indiana in June 1993, from the Research Division. In 1994, both Indiana and Pennsylvania conducted interactive editing on all samples, with Indiana implementing CAPI with seven interviewers and Pennsylvania using two CAPI interviewers. In 1995, Indiana used all interviewers for CAPI and interactive editing on all samples, while Pennsylvania and Wyoming conducted interactive editing on all samples following paper data collection. In 1996, both Indiana and Wyoming conducted interactive editing on all samples

following paper data collection. CAPI research for the June Area Frame survey was dropped after 1995. From 1994 on, the instrument was constructed to handle all versions, even though it was used only in a few states. In 1996, the instrument was still capable of CAPI but it was not used that way. The current plans for 1997 are to implement interactive editing for this survey in 10 - 26 state offices, with implementation in the remaining states in subsequent years. Blocks needed only for CAPI will be dropped.

There will be no further attempt at CAPI for the June Area Frame survey in NASS for the foreseeable future. This survey's additional hardships of using CAPI outside in bright light or bad weather along with the tight timeframe of the June Survey is too stressful for a large implementation at this time. A small research project using laptop computers for less stressful surveys is continuing.

Blaise III from Statistics Netherlands

The Blaise III system from Statistics Netherlands is used by many organizations worldwide. It is explicitly designed to integrate many of the front end survey processing needs, including data collection (CATI, CAPI, and Computer-Assisted Survey Interviewing (CASI)), interactive editing, forms-based data entry, data and survey management, tabulation, and metadata management. The recent introduction of Maniplus, a shell utility, extends the system's functionality to tabular top-down data editing and laptop CAPI management among other things.

Coordinating CAI and Interactive Editing

Even though Blaise integrates many different survey tasks into one system, the different users have their own needs and preferences. The following describes how these needs and preferences are accommodated, either naturally in Blaise, or by additional planning and work in NASS. In coordinating CAI and interactive editing, the application developer must keep in mind the needs of the users. They should also think of the optimal implementation of development standards and be prepared to change them as technology progresses.

Screen Presentation

The default screen presentations of Blaise for interviewing and editing have served well for their respective roles.

Blaise is a forms based interviewing system. Its default presentation features a split screen with question text appearing at the top and a *page* or *form* of data cells appearing at the bottom. The question displayed at the top of the screen corresponds to the position of the cursor in the *page*. Interviewers are instructed to read the question text verbatim from the top of the screen and to verify that they have typed in the correct response in the page. They can also see answers to previous questions. The forms based approach has proved to be very popular with interviewers in NASS, in part because it combines the best aspects of CAI and a paper form style of presentation.

For data editors, the default in Blaise is to use the whole screen presentation for data display. The assumption is that the editors are already familiar with question meaning and would like to see more data at one time. For most sections of the June Area Frame survey, this is a good assumption. For the more complex parts, such as the fields table, it is not. The data editor in this case can toggle the screen display

to use part of the screen for question display. Alternately, the data editor can display the question text and data definition in a pop-up window.

Question Identifiers

Different kinds of users may prefer different question identifiers. *Question names* are used in Blaise program logic to direct flow and to set edits. They also have their important uses on the Blaise screen. The interviewer needs the question name to be a *concept name*, for example, *TotalHogs* or *Wheat.Planted*. In an edit failure, it is the question name that the interviewer must choose in order to fix a problem. Thus the question name must be fully understandable in the heat of the interview. On the other hand, the data editor, usually an agricultural statistician, may prefer to see an item code as a question identifier. In NASS, the item code is a question number used for data entry. For *TotalHogs* the item code is 300. Both preferences are met by employing the concept name as the Blaise question name and the item code as the question tag. Both the tag and the question name appear on the screen next to the question's value, in the form (or page) part of the screen. A side benefit to using a concept name as the question name is that it makes for more readable code for the application developer.

Freedom of Movement

Blaise gives the data editor complete freedom to move anywhere in the instrument, regardless of routing constraints. This is especially necessary if data are collected on paper and routing rules were not applied correctly. The editor can go where necessary in order to correct or remove the data. For the interviewer, forward movement is completely controlled by the system. Backward movement is free, but the interviewer can never go off route.

Edit Message Presentation

Edit message presentation, for one edit, is the same for data editors and for interviewers. Blaise displays the edit message in an edit window, as it is written by the applications developer, and an appropriate list of possible question names and values to correct. If the edit is soft, then the user can suppress it with the <*s*> key, or fix it, whichever is needed. If the edit is hard, it must be corrected. The user can scroll in the question list. By placing the cursor on the required data item, and pressing the <*Enter*> key, the user is moved to the data item, no matter where in the instrument it is. Once the item is fixed, by pressing the <*End*> key, the user is moved to the next appropriate question.

Multiple Edits in a Form

An interviewer encounters one edit at a time, and must deal with it on the spot. The interviewer never sees more than one edit at a time. The data editor, on the other hand, has the choice of reviewing all edit messages at once, or displaying them one at a time while moving through the form. The editor can determine in which order to dispose of them and can leave them temporarily unresolved if necessary. The editor can look at all errors, just the soft ones, or just the hard ones.

Severity of Edits for CAI and Interactive Editing

There are three kinds of edit severity that Blaise keeps track of on a form-by-form basis: routing, hard edits, and soft edits. In interviewing mode, the routing is enforced dynamically. The interview is guided along the correct path as determined by the responses. In edit mode, routing is enforced pas-

sively. A routing edit marker is displayed if a cell has an answer when it should be off the route, or if it is empty when it should be answered. In edit mode, a routing edit must be cleaned up or the form is considered dirty, while in interviewing mode, the route is never violated.

Any other within-form edit can be implemented. These include additivity, relationship, consistency, linear, ratio, range, and other kinds of edits. Whether an edit is hard or soft in a Blaise instrument depends on its toggle. The key word CHECK denotes a hard edit while the key word SIGNAL denotes a soft edit. It is possible to make an edit soft in the interview and hard in data editing. For example:

```
IF CADI THEN        {data editing mode}
   CHECK            {hard edit toggle}
ELSE                {any data collection mode}
   SIGNAL           {soft edit toggle}
ENDIF
Edit statement
```

will make the edit hard for data editing and soft for interviewing. The question is when to have hard edits in the interview. Let us examine, first, an easy-to-handle edit:

```
CHECK
Planted >= Harvested
"Planted acres must be greater than or equal to Harvested acres."
```

In Blaise, the edit is stated in terms of what is correct. The edit will be invoked only if harvested acres are greater than planted. This edit is well within the interviewer's and respondent's understanding, it points definitely to an error, and it is easy to fix. Thus it is kept as a hard edit in the interview as well as in data editing mode. A more difficult edit follows:

```
Tract.Acres = Field[1].Acres + . . . + Field[99]. Acres
"Tract acres must equal the sum of field acres."
```

This edit states that the tract acres must equal the sum of all the fields within the tract, for up to 99 fields. This edit is understandable to the interviewer and the respondent, but may be difficult to correct during the interview. The required accuracy is to the tenth of an acre, and the respondent may not even know the size of the fields to the tenth. Indeed, the respondent may not know the acreage at all and thus it would be foolish to stop the interview when the respondent is incapable of giving more accurate information. This edit is soft in the interview but hard for data editing. The interviewer or data editor can use measuring devices on the aerial photo after the interview in order to determine the correct acreage of the fields. The edit should be invoked during the interview because the actual fixing of it may not be too difficult in most situations. However, for some interviews, fixing it may be very difficult, thus this edit is kept soft.

There are other edits which are hard in the traditional batch edit system. Whether they are hard or soft in the interview depends on the judgement of the developer, perhaps in consultation with other statisticians. If there is any chance that a hard edit can in truth be violated, then it is softened during the interview. In the June Area Frame survey, most edits which are hard in the traditional batch edit are hard in the interview.

Don't Know and Refusal

There are some questions that farmers are not able to answer or refuse to answer. The CAPI instrument allows for *Don't Know* and *Refusal* for selected questions. For example, the total land question can take the *Don't Know* response in CAPI. The summary system, on the other hand, expects important items to be filled in. The method of imputation is hand imputation by the editor during the edit process. Sources of information available to the data editor include the respondent's previous survey forms if any, and since this is an area frame survey, interviewer observations from the road. It is felt that manual imputation of observed values, or values from previous years, is more accurate than automated imputation. The result of this is that questions are not allowed to have a *Don't Know* or *Refusal* response in edit mode and thus there is a subtle difference in data definition between the two modes. The developer has to put in edits to flag these nonresponse codes in the edit.

Zero and Empty

In data collection, according to NASS standards, an *EMPTY* cell on the route represents a question that has not been asked yet. The interviewer must enter a number, even if it is zero. In paper data collection, and in data editing, an *EMPTY* cell represents a zero. This represents another subtle difference in data definition between modes of use of a Blaise instrument in NASS. The instrument has edits that allow the *EMPTY* in data editing, but not in data collection. An enhancement NASS would like to see for its dual use instruments is a key word such as *CADIEMPTY* that would be written as part of data definition. This would tell the system to allow *EMPTY* for data editing, but require a response for interviewing.

Screening Questions

On paper forms, and in the CAPI instrument, there are screening questions at the start of each section. For example, in the hogs section, questions are asked to determine if there are hogs on the farm. If not, then the rest of the section is skipped. However, when data are collected on paper forms, the screening questions are not key punched in NASS. This is done to save the data entry people a few key strokes. Routing in CAPI is dependent on the values of the screening questions as entered by the interviewer. Routing in data editing mode is dependent upon the presence and values of the commodity items themselves. In this latter situation, it is up to the instrument to determine, from the values in the section, what the values of the screening questions should be, and impute them. NASS should consider key punching the screening questions.

Switching Between Modes

Blaise can allow the user to switch between its four modes of operation. It also allows the application developer to disable this feature. The mode switching is disabled for interviewers but is enabled for data editors. The editors may find it advantageous to switch to the interviewing mode to follow the course of an interview.

CAI Administration Blocks

CAI administration blocks include blocks of questions that are necessary to administer a CAI questionnaire. As implemented in the June Area Frame Survey, these blocks include a *Name and Address* block for the respondent and partners, an *Appointment* block, a *Nonresponse* block, and a *Time and*

Date Stamp block. These blocks have marginal use in data editing, but in the multi-mode nature of Blaise, they are present in edit mode as well as data collection mode. This takes up space in the data set and increases the number of screens it takes to present the data. This has not been much of a problem in practice. The developers give the data editors jumping points that get them to the commodity data quickly, and save them the trouble of paging through many administration screens. Navigation through edit messages in Blaise is superb and gives another possibility to get to the problems without paging. It is also possible to hide the administration screens in edit mode, but this has not been done.

Mouse and Operating System

Blaise III allows functions to be executed either by mouse or by function keys. The mouse is turned off for interviewers but left on for data editors. The editor has a choice of which to use.

Blaise III is DOS-based, but can be operated in a DOS window of Windows 3.1, Windows 95, or Windows NT. The interviewers are generally set up in a straight DOS environment. Their job is to conduct interviews, they do not need to switch between tasks. The data editors usually operate out of a DOS window in Windows 3.1. The data editors in NASS are agricultural statisticians with a wide variety of duties to carry out. When the Blaise edit is in a DOS windows, the data editor can switch to other tasks without having to close the Blaise application.

Data Read-In from Paper Forms

When data are collected on paper, it is necessary to read them into the editing instrument. If data are collected only in CAPI, then edited in the same software system, then no data transfer is necessary. Data entry in NASS is done in another system. The kind of data entry is known as item-code data entry, where the 3 digit question number is entered followed by its value. The data are put out into 80 column records where many physical records represent one logical record. The Manipula utility is used to read data into Blaise format. For each data item, a mapping from the Blaise record to an item code must be made. This mapping is easily accomplished by using the item code as the question tag for most of the June Area Frame instrument. Cameleon, a meta-data utility, generates the necessary Manipula program to do the data read-in. However, for the grain stocks table and the crops table, much of the metadata are held in an external Blaise file. A separate Manipula program is required to read these data in. Since the meta data in this case are not part of an instrument's code, Cameleon cannot generate this Manipula program. This part of the data read-in is accomplished by using the external files to mediate the meaning of each data item for each state.

Farmer in More than One Segment

A data collection challenge is that a farm respondent may operate land in two or more segments. However, the traditional unit of processing is based on, and the paper questionnaire represents, a tract. The questionnaire has farm-level questions and tract-level questions. When data are collected on paper, and the interviewer finds out that the respondent operates land in two or more segments, she collects farm-level questions on one just form. For the tract-level questions, she shifts back and forth between paper forms.

Instantly shifting between forms was not so easy to do in CAPI in Blaise in 1995. One method of handling a respondent in 2 segments that was tried in 1995 was to have two instances of all tract-level blocks in the instrument, including the massive field table (each instance of which had 5,500 possible

questions). Mechanically this worked in the interview, and was easily implemented by the instrument authors, but was not optimal for the edit. This scheme greatly enlarged the instrument in terms of numbers of questions, edits, and screens for a situation that rarely occurred. It slowed down batch processes such as combining forms from different CAPI machines into one data set. Another problem with this approach is that there is no upper limit on the number of segments that a farmer might appear in. As soon as two instances of tract-level questions were allowed, someone might find a farmer who operated land in 3 or more segments. If data for two tracts were collected in one CAPI form, there was the problem of transferring data from one form to a second. In future, this difficult data collection problem would be handled through the use of the new shell utility Maniplus. This utility makes it far easier to shift between forms for tract-level questions for one farmer in multiple segments, and then to automatically transfer farm-level data from one form to the rest.

Benefits of One System for CAI and Interactive Editing

While the above paragraphs enumerate some differences between implementation of editing and interviewing in one system, there are powerful reasons for the adoption of one system for both tasks. Despite the differences that must be accounted for, 80 percent or more of the program code is applicable between the two modes of processing, including data definition, routing, edits, and data base structure. The percent of shareable code would be higher if NASS were to adopt different conventions to eliminate the difference in data definition between the two modes. Though there is more programming code than for CAPI alone or for interactive editing alone, it is still far less code than if two programs were written in two systems. Where there are two systems, additional conversion code must be written or generated to transfer data from CAPI to edit. Data conversion itself can be problematic between two different systems. Conventions for handling interviewer remarks or nonresponse codes such as *Don't Know* or *Refusal* may be incompatible, or may require a lot of work to bridge. Database structure for a complex multi-level instrument such as the June Area Frame survey may only be handled adequately in a data collection system.

With everything in one program code, the developers are forced to make explicit methodological choices between what happens in data collection and what happens in edit. If there are two systems, sometimes these choices are made implicitly without anyone realizing it.

The Implementation of Top-Down Edit Methods

NASS for some time has had paper-based top-down editing tools. These are SAS programs that generate various kinds of data listings, tables, and other macro views of the data that allow the agricultural statistician in the state to quickly spot problems. Recently, a SAS-based top-down system has implemented most of these ideas in an interactive mode for the Hog Report. This system, called IDAS (Interactive Data Analysis System) was demonstrated at this conference (Hood, 1996). For the Hog Report, it is the intention of NASS to implement this system in 30 quarterly states and to retain most of the SAS paper-based reviews for other states which do the survey only annually. NASS (Knopf et al., 1996) has just embarked on a systematic review of the overall Hog Report editing process, from data collection, through a micro-level interactive edit, and the macro-edit tools. Knopf and colleagues have made explicit decisions about which edits to invoke at each stage of the process. Some edits have been eliminated or consolidated. A few of the range edits are no longer implemented in the micro-level, post-collection, interactive review, but are invoked in CATI and at the macro-level.

A new utility in Blaise, Maniplus, is recently available. Things that can be done with it include the production of a laptop computer CAPI management system, a data and survey management system for the in-office processing of forms, and tabular top-down editing methods. The latter function has been demonstrated in the form of interactive data listings though there is no limitation on the form of tabular review. It is possible to scroll through a list of records, sort them in any order, invoke a form, change the data, get back into the interactive table, and immediately see the effect of the changes on the table and on estimates. Maniplus does not have a natural graphical review capability as the IDAS system does. On the other hand, Maniplus knows the metadata of Blaise instruments, and can read survey data directly from the Blaise database. This allows for top-down review of data without having to convert data to a different system. NASS should consider its use for interactive data listings, getting away from the paper-based data listings it plans to continue.

References

Knopf, D.; Anderson, C.; Apodaca, M.; Pallesen, M.; Prusacki, J.; and Tesky, M. (1996). Report of the Hog Editing and Analysis Team, internal NASS report.

Hood, R. (1996). Improving the Quality of Survey Data Through an Interactive Data Analysis System, demonstration at the *Data Editing Workshop and Exposition*, Washington, DC. ■

Note: The work represented in this paper was done while the author worked for the National Agricultural Statistics Service (NASS), U. S. Department of Agriculture. The views and conclusions are the author's, not those of NASS or Westat, Inc. The author thanks several people in NASS and Westat, Inc. for reviewing and commenting on a draft of this paper.