ONE AGENCY'S EXPERIENCE WITH THE BLAISE INTEGRATED SURVEY PROCESSING SYSTEM

Mark Pierzchala, National Agricultural Statistics Service (NASS)
Room 4168 South Building, USDA, Washington, DC 20250-2000

Key Words: Editing, CATI, CAPI, Data Entry, Survey Management

Abstract

NASS has been testing and using the Blaise survey processing system from the Netherlands Central Bureau of Statistics for three years. Over 20 surveys have been programmed in Blaise, more than half of them in the state offices. The two largest instruments constructed in Blaise have been a prototype CAPI instrument of 1,200 questions with a relatively complicated structure and an editing instrument of 4,000 questions with a relatively simple structure (question number is independent of the number of edits, etc.). Two successful CATI applications of about 4,000 records were conducted in mid-1991 using the system. Blaise combines great power with ease of programming and has been programmed by people of varying programming backgrounds with little or no training. The Blaise system can handle all of NASS's surveys rather easily and can be used in all locations. In order to standardize the use of Blaise in NASS's 46 locations, standards have been formulated, a library of Blaise code has been established, and an applications manual published for use in NASS. Blaise has been designated as the official interactive editing system in NASS.

About NASS

NASS employs about 1000 people of which 3/4 work in 45 state offices while the rest work in headquarters in Washington DC. For national level surveys, work is carried out in the state offices under the direction of headquarters. Data are hand treated in each state office and processed on one mainframe computer located in Florida. NASS is procuring a Local Area Network (LAN) for each of its locations. When the installations are complete there will be enough computer power in each office to transfer all of NASS's day-to-day processing from the mainframe to the states. As the processing is shifted from leased facilities to owned LANs it will be possible to adopt new processing techniques, for example, interactive editing of survey data.

One implication of transferring data processing from the mainframe to the 45 LANs is that it will no longer be possible for headquarters to technologically control the survey process through one mainframe, thus headquarters personnel worry about losing control. As a result, some people would prefer to wait for better communications between headquarters and the LANs before implementing interactive editing.

In addition to the federal level surveys, NASS's state offices also conduct many state-level surveys. Formerly these surveys have been edited in a wide variety of software on either the mainframe or on microcomputers. For these surveys, the state personnel do not get much if any assistance from headquarters. Thus they must have a system that is easy to use and program.

Interactive Editing Working Group

On the national level, the author and others in NASS felt that the agency should move to interactive editing which would be possible on the LANs. An Interactive Editing Working Group (IEWG) was formed in May 1989 to investigate how interactive editing might be done. The group planned to do three surveys by the end of 1989. These included the Corn Objective Yield (COY) survey, the Agricultural Survey program (September and December) and the survey of Prices Received for Livestock.

Corn Objective Yield Survey

The Corn Objective Yield survey is conducted monthly from July to December (five months) in corn fields in 10 states. It has a small sample size but is collected on five different forms, any one of which (or more) may be used in any month depending on the progress of the corn field. In 1989, just one of the five forms (the B form) was programmed in Blaise while in 1990, all five forms were included in the program. The B form is the most complicated of all the forms and holds about 60 percent of all the information collected. The challenge in going from one form in 1989 to 5 forms in 1990 is that in the latter year it was necessary to take into consideration the complex survey management between the five forms over five months.

Programming for the COY was complicated by two major impediments: 1) the power of the Blaise system is not well documented, and 2) some unnecessary conditions were put on the COY test by other units in NASS. Blaise has many powerful programming features such as macros (and nested macros). These are not found in the documentation. As a result, many chunks of code were programmed two, four, or eight times when they could have been programmed once. In addition, the regular

COY edit is programmed in SAS, and it was required that all of the SAS edits be duplicated as exactly as possible in the Blaise COY edit. This meant that some edits which could have been programmed as Blaise route statements (an easy thing to do) were programmed instead as regular edits (somewhat more difficult since essentially the edit is describing routes and inappropriate routes). As a result, the code for the COY instrument in 1990 was about 2 to 3 times longer than needed. In addition, some aspects of the COY are not very well documented and some of the survey management so obscure that it took a lot of extremely tedious work just to find out what needed to be programmed.

Nevertheless, an instrument was built that took care of all editing, survey management, and progress reports for all forms and for all months. Despite some initial difficulties, evaluations from the field offices were very positive and recommended that Blaise replace the SAS batch edit. The editors were able to receive instant feedback each time they pressed the edit key and knew when the form was clean. Also, since they were editing on their own computers, they could edit whenever they wanted and not have to wait for an overnight job to be run. Blaise took care of all survey management and the related tabulation program Abacus was used to generate about 40 different reports most of which were never used.

Some improvements in the Blaise system that would have eased the programming would have been a system feature for generating error counts, and some robustness provisions. Error counts were generated, but these were specially programmed. As regards robustness, it was necessary to program some extra cells for contingencies and also some extra edits. Once these extras were in place, then the instrument became very easy to change in the middle of a survey.

December 1989 Agricultural Survey

The Agricultural Surveys are conducted quarterly, are medium sized (about 200 questions per state) and have a larger sample size (about 2000 per state). It is the largest survey that NASS does in a two week survey period that can feasibly be edited twice at one time (once in batch and once interactively) in order to compare the two processes. The December Agriculture survey is the third or fourth largest survey that NASS conducts on a regular schedule.

Interactive Editing After CATI

The Ohio office collects data for the December Agricultural Survey through CATI using the CASES system from the University of California at Berkeley. Normally, the data are then fed into the mainframe edit and then cleaned up on batch printouts. In December 1989, the Blaise edit was inserted between the CATI collection and the mainframe edit.

The data editors liked the Blaise edit because they had a form (on the screen) to look at when editing and not just a paper printout of selected values and error messages. There was no parallel control process in this editing test, however the Ohio data editors felt that the quality of the whole process improved over the normal batch edit.

Interactive versus Batch Editing

In Wisconsin the data were collected on paper. Data were edited in many different modes in the interactive trial and were also edited in batch by the regular state personnel. For the interactive edit, people travelled to Wisconsin from headquarters. For the most part data were edited first in Blaise without a previous hand edit and then the paper forms were given to the state people to do what they normally do. The trial took two weeks thus some of the interactive editors were only able to participate for one week and had to be replaced. The trial was viewed as a "proof of concept" of interactive editing and cannot be considered as having been conducted as a true experiment. Various time and productivity measurements were taken for the various interactive methods used as well as the batch control. Also, expansions were generated from both processes. The Blaise interactive edit was conducted in 20 percent less time than the regular paper batch edit. Expansions from the two processes were very similar for almost all of over 70 items. A few of the expansions differed considerably. Extensive follow-up revealed that the few differences in expansions were due either to editor differences or to missing edits in the Blaise instrument in one section.

Prices Received for Livestock

The survey of Prices Received for Livestock is conducted monthly in livestock auction barns. Its distinguishing feature is the large repetitive tables that hold the price information. The tables are wider and longer than the microcomputer screen. Blaise has some powerful tabling features and was able to handle this survey without problem. It was taken to the Nebraska state office and worked well there. As this survey was absolutely no challenge to the Blaise system, further research work on it was dropped.

Recommendations of the Interactive Editing Working Group

After the program of the Interactive Editing Working Group (IEWG) was completed in December 1989 the group met in January 1990 and published a report in February 1990 which made three major recommendations: 1) that interactive editing be adopted by the agency, 2) that Blaise be the interactive editing system, and 3) that Blaise be evaluated as a data collection software. The recommendations were made subject to some conditions:

1) that it be demonstrated that Blaise could handle all of the agency's surveys and 2) that an adequate working relationship be developed between NASS and the Netherlands Central Bureau of Statistics. The recommendations that Blaise be adopted as NASS's operational editing system and be evaluated as a data collection language immediately ignited a storm of protest in NASS, mainly coming from those who work with the CASES system from the University of California at Berkeley. They felt that CASES should be the agency's interactive editing system (even though this capability is not available in CASES). As a result of the recommendations of the IEWG and the protests of the others, a new group has been formed in NASS called the Interactive Survey Software Committee (ISSC). This committee has been formed to take a look at all of the interactive issues relating to editing, data collection, data entry, and their integration and to make suggestions for evaluating both systems in all three areas. The work of the ISSC is still proceeding.

Other Work With Blaise

NASS's experience with the Blaise system has not been confined to the work of the IEWG. Nationally, questionnaires or parts of questionnaires have been programmed in Blaise from February 1990 to the present for editing. As well, Blaise has been used in two CAPI experiments and in two CATI tests.

Most of the work in 1990 was dedicated to proving that Blaise could handle all of NASS's surveys. This meant that it had to handle the size of NASS's largest surveys as well as their structure. The largest surveys in NASS are the Farm Costs and Returns Survey (FCRS) which comes in about five different versions each year, the Agricultural Chemical Use Survey (pesticides and fertilizers), the June Area Frame survey, and the December Agricultural survey (treated above).

Farm Costs and Returns Survey in CAPI

In early 1990, a CAPI instrument for the FCRS expenditures questionnaire (1200 questions) was developed in Blaise. A few interviews were conducted in state offices. In early 1991 the exercise was repeated which wasn't trivial as much of the questionnaire was changed from 1990 to 1991. In 1991, some farmers (not actually in the sample) were interviewed by enumerators who had no training either in CAPI in general or in Blaise. The major benefits from doing the FCRS were that 1) it was a great training tool for learning how to program in Blaise, 2) learning that Blaise could handle this formidable survey, 3) learning that enumerators could easily learn to use the Blaise system. On the negative side, it was found that Blaise produced an executable instrument that was very large and that without training the enumerators tended to look too much at the bottom part of the screen (the forms part) and not at the actual question text in the top part of the screen (see illustrations, last page). Programming was carried out both years on a spare time basis by three or four people. That is, in both years it was an unbudgeted project. The programmers look forward to doing this survey in the future when more time and money will be devoted to it.

Operational Editing in Blaise

On the National level, two surveys have been programmed in the editing mode. These are the Aquaculture Survey and Cotton Ginnings Survey. These are surveys that people did not want to do (or could not do) in other NASS editing systems.

Aquaculture Survey

The Aquaculture Survey is conducted quarterly in four states and semi-annually in 17 states. There are about 100 questions (cells) in the instrument. The first quarter's processing was done in April 1991. The four states involved had never processed data in Blaise before. Training for the data editors was done through a tutorial. The April processing proceeded smoothly and the editors said that they liked using the Blaise system. Since then, a larger instrument was constructed that handles all four quarters. In July the instrument was used in all 17 states. Though the survey is not conducted in CATI at the moment, the Blaise instrument was constructed to switch to that mode as well in the future if necessary (with some more work and testing to be done). The Aquaculture instrument features three tables each of which can be completed in different ways. The challenge in the data collection mode is to allow the enumerator some freedom of movement in the tables while at the same time enforcing certain completion criteria. At the same time, there are some things that are required to be done in the editing mode but not in the data collection mode and vice versa.

Cotton Ginnings Survey

Cotton Ginnings has just been programmed in Blaise as an editing instrument. This is a small survey collected over 13 periods within a 6 month period. While the amount of data collected is small, the survey management between the survey periods is complex. Each state is left to collect and manage the survey as it sees fit, however the editing is standardized through the Blaise instrument.The instrument must be capable of giving information to the states in the various ways that it is needed.

In developing and planning many NASS BLAISE STANDARDS were broken. For example, specifications for the instrument were still being received in mid-October whereas the survey was being conducted from the

first of August. This occurred because the survey is new to NASS, the survey managers did not know what Blaise could do, and because states have made suggestions for improvement. Thus there has been a cycle of specifications that has taken place during the survey.

On the other hand NASS BLAISE STANDARDS for programming and testing have been met. The instrument was designed to accommodate major enhancements during the survey (the programmers knew that more things would be coming, they just did not know exactly what they would be). The instrument has gone through five revision cycles and placed on the LANs in the state offices five times without interrupting the survey.

June Area Frame Survey

The June Area Frame Survey features a crop table 48 rows by 100 columns in addition to about 200 other questions making an instrument of about 5000 potential questions altogether (number of questions is independent of edits, auxiliary calculations and the like). The crops table was constructed and put together with about half of the FCRS instrument. Probably in operational use, the actual size of the table would be reduced as it takes several screens to represent it both width-wise and length-wise. However, the exercise proved that the Blaise system could handle an instrument far larger than the June Area Frame Survey.

Agricultural Chemical Use Survey

Another massive survey in NASS is the new Agricultural Chemical Use survey which is composed of several related data collection efforts. This survey collects information on crops and the pesticides and fertilizers used on them. This information is collected on table after table of similar structures but different topic matter. For example, the tables for herbicides are identical in structure to those for insecticides, however the entries are compared against different lists of edit limits. The Chemical Use Survey features a kind of rostering where a first table of crops will be filled in and subsequent tables of lines of chemical use information will be compared against the first. However, in this survey, the row entries in the second set of tables can be filled in any order. Thus it is necessary to compare the second table line entries which are entered randomly to ordered entries in the first table. The second challenge of this survey is comparing rows of chemical usage information against large lists of chemicals and edit limits, line by line. The lists of edit limits can be a few thousand lines deep and may be accessed by up to three or four indices (e.g., crop x pesticide x formulation x pest). The third challenge is accumulating information for similar lines within a table when these lines may not be one after another. For example, the same pesticide may have been applied to a crop two or more times. It may be necessary to find these similar lines and add up their rates of application to determine if the sum of rates violate edit limits. The fourth challenge is that on some of the larger farms, the amount of collected data may become very large. (The collection effort on paper took several days on one farm.) Thus contingencies must be built in to the instrument to be able to handle the large amounts of data per sample. The table and macro structure of Blaise make it possible to meet these challenges fairly easily. However, the coding of the instrument must be done in a certain way in order to ensure that the performance of the instrument does not deteriorate as massive amounts of data are collected.

CATI

Two CATI applications were conducted in Blaise in early summer 1991. More testing should be done but no further tests are currently scheduled.

Criteria Letter in Ohio

The Ohio office conducted a CATI survey for a criteria letter for list building (about 100 questions) for which they programmed the instrument in Blaise with some review of code from headquarters. About 3,000 calls were made. Enumerators were able to learn the system easily and liked using it. (Ohio has been using CASES for CATI for several years.) The Ohio test proved that enumerators could be trained to read the text from the question part of the screen and to not look too much at the forms part of the screen (which had been a concern with untrained enumerators in the FCRS test).

Cattle On Feed in Iowa

The Cattle on Feed Survey in Iowa was used to compare the Blaise system to the CASES system for data collection. In Iowa, CASES has been used for some time as the operational CATI system, thus special training for Blaise was conducted. The Blaise CATI call scheduler and manager were used for the Blaise calls and was received favorably by the office CATI coordinator and enumerators. (NASS does not normally use the CASES call scheduler). About 900 calls were completed with both systems. The enumerators were asked to state a preference for a system. Of 18 enumerators, 10 chose Blaise, 5 chose CASES, and 3 had no choice. Enumerators were asked to rate Blaise against CASES on 14 criteria such as screen appearance and ease of use. Blaise was rated above CASES on all but one criterion. The Blaise instrument required about 60% less code when measured in numbers of sheets of paper of code (16 pages vs. 42).

State Office Use of Blaise

An important aspect of any system meant for use on NASS's LANs is the ease with which state personnel can

create their own applications. The state offices are often asked to conduct state-level or regional surveys. These surveys are often conducted in more than one state over the years. The introduction of Blaise gives each state office the opportunity to develop editing applications in a standard software. Once written, these applications can be shared between offices. The receiving offices can then modify already created instruments rather than starting from scratch. About 15 instruments have been created in state offices that have served as trial centers. Wisconsin has programmed applications for Mink, Fruit Tree, Ornamental Nursery, Pesticide, Dairy, and Crop Disposition surveys. Virginia has programmed applications for Soybean Variety, Apple Disposition, and other surveys. Pennsylvania has programmed applications in Turf Grass and Equine Surveys. Ohio has written an instrument for collecting list frame information in the CATI mode.

State Office Programmers

Most of the instruments in the state offices have been written by people who have received little or no training. For the most part, the state level surveys are small but on occasion they can be quite large and complicated. For example the Fruit Tree survey in Wisconsin contained over 4,000 cells.

Standards and Applications Manual

In order to facilitate the writing of NASS-type applications in state offices (as well as in headquarters), a NASS Blaise Applications Manual for programming in Blaise has been written. It contains a programming tutorial. An integral part of the manual are NASS BLAISE STANDARDS that cover all aspects of conducting a survey in Blaise. Given that the standards are followed in the states, then it become possible for states to share their applications efficiently. In addition to the applications manual, a library of code has been started. For most surveys, the states will be able to grab significant chunks of code from the library. As state instruments are written, they will be checked for adherence to standards and placed in the official Blaise library. In the past few months since the distribution of the software and manual, several people have worked through the tutorials included or have started to do so (25 people have signed up for Blaise training in headquarters).

Findings

NASS has found Blaise to be an easy-to-use, relatively easy-to-program system with a wide range of capabilities. It can handle all of NASS's questionnaires as regards size and structure. Most of NASS's work has been in the area of editing, but also some trials have been conducted for CATI and CAPI applications. On occasion, Blaise has also been used for data entry, though usually data have been read into Blaise from other data entry packages. One of the features of Blaise that will make it very easy to disseminate in NASS's 45 state offices is the fact that the system has a menu-accessed development environment. For example, the system comes with a powerful and easy-to-use text editor and a wide range of utilities. One of the outstanding utilities is the setup generator that can produce dataset (data step) code in a wide variety of languages such as SAS, SPSS, ORACLE, STATA, etc., and can be programmed to produce setups for in-house languages.

People of widely varying backgrounds have been able to program in Blaise quite easily for editing applications. However, people tend not to use all of the considerable power of Blaise unless it is pointed out to them. In the author's opinion, the Blaise system combines great power with ease of programming. Also the rate of improvement of the system has been very good. NASS has not required a great deal of support from the Netherlands Central Bureau of Statistics and has never had any training from the CBS.

Improvements

There are several improvements that could be made to the Blaise system. For editing, it would be nice to have an error count feature when data are edited in the batch mode. In addition for both the editing and data collection mode, the system could use an audit trail (though it is possible to program one). Some improvements in the programming language would be helpful. For example, Blaise codes **Don't Know** and **Refusal** as numeric values just outside the valid range of the question. The awkward part of this convention is that when performing calculations, the developer must take pains to screen out questions with these responses. The Blaise language should automatically screen out theses values. For CATI, some text enhancements would be nice though not strictly speaking necessary (in the Iowa CATI test, the CASES instrument used text enhancements, the Blaise instrument did not). (This is a point of contention in NASS. For example, CAPI applications will be collected on computers with monochrome screens for a long time. However, some people state that some text attributes are absolutely necessary, especially color. Text enhancements include the ability to use color, bolding, reverse video and the like.) In the data collection mode, there could be more flexibility in the display of error messages. It is possible to make free-form comments in a separate text file, however once stored the message cannot appear back on the screen (it can be printed). It is NASS's understanding that these features in some form are being added to the next major versions of the system.

The Netherlands CBS has a solid track record of

coming up with imaginative and appropriate solutions in enhancing the system. The reason for this, in the author's opinion, is that Blaise is extensively used in all modes in the CBS. As a result, almost all problems are already well known, and solutions suggested by CBS personnel many months before outside users run across the same problems. NASS offers few original suggestions to the CBS.

The Blaise documentation in English is very good. Currently there are 12 manuals for the various parts of the system including manuals for associated software packages or utilities. The last manual out is called **Tutorial** but is more like an applications manual. It is very good but needs to go further yet to show how to organize the coding for large and complicated instruments.

Procurement of Blaise

NASS has purchased a license for all of the 45 state offices and headquarters units. In addition, a separate support agreement has been put into effect. Blaise will be a standard data processing tool among several other tools available on the state office LANs and in headquarters. The extent of future use in NASS has yet to be decided.

FIGURE 1 A Blaise Editing Screen

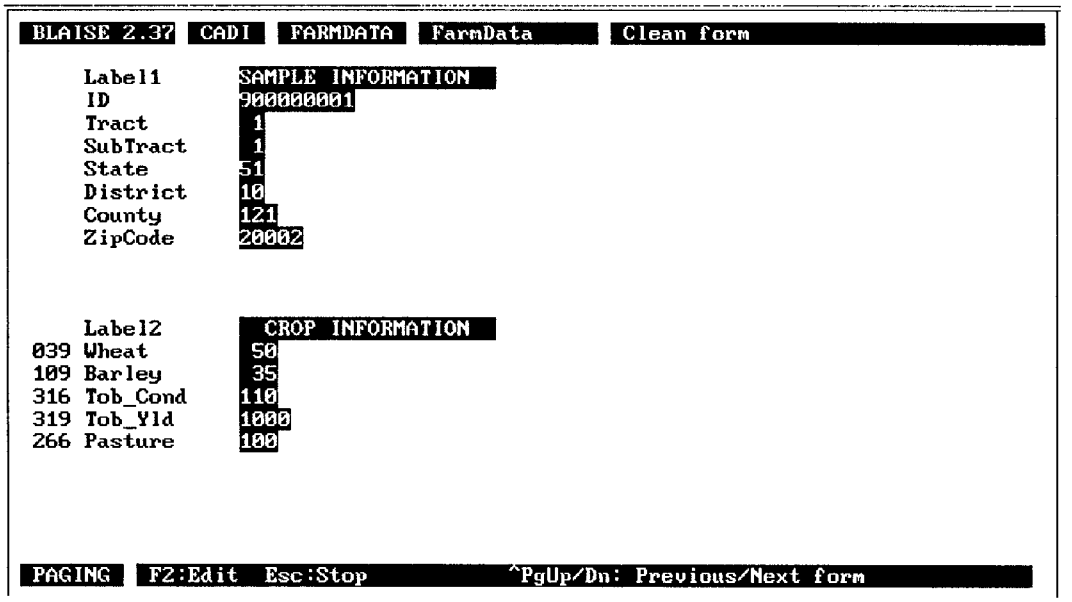

FIGURE 2 A Blaise Data Collection Screen