A Report of Procedures Used in a Circulation Survey of a Public Library

by William F. Morse and E. Walfred Erickson

Graduate Students, University of Illinois Library School, Urbana, Illinois

A circulation survey was made in November 1950 as part of a long-range study of the Urbana (Illinois) Free Library by the University of Illinois Library School. The Urbana Free Library serves a community of 23,000 people and has book resources of approximately 40,000 volumes. With limited finances, personnel, and time available to undertake such a project, a method of conducting the survey had to be devised which would produce relatively complete and reliable data. This is a report of that survey, and of some of the lessons learned from it. The system used to record the information consisted of mark-sensing cards which were punched and tabulated by International Business Machines. The success of the method depended on accurate codes which would secure all desired information and produce tabulated results in a form conducive to interpretation and analysis. Equally important to the success of the survey was a method of detection of error that would hold both human and mechanical error to a minimum.

The scope of the survey covered everything that could be learned from existing records about every circulation for a period of thirteen days, November 8-21. This period was one-half of the number of days the library was open during the month of November 1950. November represented 8.5% of the year's circulation, 0.1% above the average monthly circulation of 8.4%. The basic information sought was knowledge of each item circulated during the period tested. This included (a) information on the type of material borrowed, (b) previous circulations of the book as revealed by the book card, (c) information concerning the book, viz., number of pages, date of publication, date of acquisition, or, in the case of magazines, the year of publication, (d) the borrower and his family, e.g., juvenile or adult, sex, member of a family unit, occupation, and residence, and (e) length of retention of material, including renewals and overdues.

This information was recorded directly on IBM cards rather than on mimeographed work sheets; the amount of time used to record such data for the 4,094 items circulated was approximately 812 hours. The cards used are known as mark-sensing cards because they were marked with a soft lead pencil and then these marks were translated into punches by a machine process based on the sensitivity of an electric current to the carbon in the lead. Each card consists of 27 columns per side with spaces 0-9 in each column, each column representing a piece of coded information. Only one space in each column can be marked if correct information is to be tabulated by the International Business Machines. For instance, to record the pagination of a 250 page book, three columns were marked--in the first column, space number 2; in the second, space number 5; and in the third, space number 0. With 27 columns on each side of the card there were 54 columns in all that could be marked. However, the machines could punch a total of 80 columns on the card, 26 more than could be marked with the soft pencils. These extra columns were used for the coding of information that could be punched directly by the machine without marking them first (this is called "gang punching"), e.g., the date borrowed, which could be punched at one time in all cards for books circulating in a given day.
Errors were revealed by three means: (a) by having one individual spot check a sample of cards, (b) by the machines while punching the cards and through internal checks, and (c) by various individuals working with the survey who encountered and corrected errors incidentally in the course of their work. In our case, where the desired information was divided into five major parts, the spot-checking of the cards revealed error rates of 1%, 1.2%, 2.3%, 0.4%, and 1.4% respectively for the five sections. Not only did this machine method produce relatively complete and reliable data, but it also leaves us, as a permanent record of the two week period, a simple file of cards, each containing all relevant information on a given circulation transaction. These cards can be sent through the International Business Machines for further cross-tabulations.

The Use of Codes

Certain essential procedures and time-saving devices useful in the process of recording information by this method were discovered in conducting this survey. First, a complete investigation of the charging system, registration files, and library book records (such as the shelf list, accession record, and library catalog) should be made. With a full understanding of the types of information available from each library record, one can then proceed to devise codes. In compiling codes it is desirable to utilize every space in every column economically, being careful to avoid conflicting data. For example, in the case of books, the same columns for the Dewey decimal numbers were used for the fiction title code because no one book could be both fiction and non-fiction. The same thing held true for the coded list of periodical titles, which occupies the same columns as the Dewey decimal numbers and the fiction code. Much space can be saved by pairing such codes, after inspecting all items of information and ascertaining which are mutually exclusive.

Information necessary for easy identification of punched cards in the various stages of the survey was written in ink at the top of each card. This included the call number, the title (in the case of unclassified materials), the accession number, the date the book card was typed, and the initials of the marker. This had to be done in ink because the brushes of the punching device in the machine are sensitive only to the bold mark of a lead pencil and will not be attracted to ink on the card. This identifying information was essential to the process of marking data directly onto the permanent cards (instead of using mimeographed work-sheets) in several different stages.

Although much of the data desired for an analysis of public library circulation seems at first glance to defy coding, with careful study of the problem even complex information can be made to conform to the limitations of the IBM card. This can be illustrated by the codes devised in this survey to record the number of borrowers within a family group who borrowed books during the time tested. First, a borrower's library registration card number was recorded for each individual; then a family number was assigned all borrowers. This number was the first registration number in the library's alphabetical registration file for any person of the given family name and address. Hence, in some cases the same borrower's number is recorded on an IBM card twice, once to identify the borrower and once to identify his family. This device makes it possible to tabulate information relating to all members of a given family who used the library in this period.

In the case of a library the size of that of Urbana, with a circulation of 92,818 for 1950 and run by a small staff (4 full-time persons), the use of their records must be designed to cause the minimum inconvenience to the daily operation of the library. This necessitated a procedure to obtain information quickly each day from the book cards for materials which had been loaned the previous day. Furthermore, to obtain information on the length of time an item was retained by the borrower, a thin slip of paper (with sufficient information to identify the item) was stapled to the
book card at the same time that information concerning the book's circulation was recorded on the IBM mark-sensing card. The book cards were then returned to the library's circulation file. When the book was returned, the date was stamped on the slip of paper which was then detached from the book card. Later this information was recorded on the IBM card to provide information as to how long different kinds of people keep out different kinds of books.

To utilize fully all the possibilities of the IBM card in recording all available data, it was often necessary to devise supplementary codes within the framework of the basic code. In the Urbana survey it was found necessary to form such codes for the place of residence of a borrower, the occupation of the borrower, the list of periodical titles, and the authors and titles of fiction books. The residence code consisted of assigning numbers to every block in the city on a large map, so that the street address taken from the library's registration file could be assigned a block number. This will reveal information concerning the distance of residence of borrowers from the library and its relationship to the use of the library. The fiction code involved alphabetizing the cards and assigning consecutive numbers to the titles borrowed. The occupations' code was based on that used for registration of borrowers in the Gary Public Library. A list of periodicals received by the Urbana Free Library was checked against Ulrich's Periodical Directory, and those titles found listed were arranged alphabetically and assigned serial numbers.

Perhaps the most important procedure in conducting a survey using this method of recording and tabulating information is the pre-testing of the codes. The pre-tests serve to check for fundamental errors in the codes, which if not found and corrected, could negate the entire survey. The pre-tests also form the basis for designing the pattern for people recording the data. This pattern must consider the library's daily needs for its records, the logical use of them for recording consistent information, and the most economical use of manpower. For instance, it was found that the recording of the Urbana survey data could be divided into five parts. First, all materials to be recorded from the book card had to be done daily in order to release the cards for the library's use. Second, items which would produce information about the borrower (viz., registration files, directories, occupational codes and handbooks, and a city map) contained interlocking information which called for small teams and efficient arrangement of the materials to speed the flow of information to be recorded. Third, information concerning the book itself was drawn from the shelf list and other bibliographical guides (such as Cumulative Book Index) and the work had to be arranged in a manner consistent with the effective use of these tools. Fourth, the alphabetizing of the fiction cards and the separation of the repeat circulations and duplicate copies for coding were done in a separate process by a few people while the non-fiction cards were being punched. Finally, the information regarding the length of retention could be recorded any time after the return of the books and before final tabulation of the data by the International Business Machines, as this information was not dependent on any other recorded.

The pre-tests also indicated the maximum and minimum speed it would take to mark the IBM cards. This information made possible the setting up of work schedules by which available help could be used most economically and effectively. For instance, teams of a dozen or more persons were needed to record information from the daily circulation to release that record to the library as soon as possible. In the case of borrower information it was found that teams of more than three people could not work well with the limited number of tools and available space. Since the workers in the survey varied in experience and background, ranging from that of professional librarians to students who knew practically nothing about librarianship, it was essential that detailed instructions be drawn up if uniformity in marking was to be achieved. These instructions, in addition to basic details, included interpretations of irregularities
Checking for Error

On the premise that a survey is only as good as the reliability of the data it produces, it was decided early in the planning that procedures for the detection of error would function concurrently with the marking of the cards. By a systematic check for error it was hoped that the quality of the marking would be improved, the rate of error would be held to a minimum, and the type of errors being made would be ascertained. It was decided that one person would make all checks for error, e.g., it was his duty to spot-check the work of the markers each day. Using a sample of 10% of all cards marked, the third card and every tenth card thereafter of each marker were checked. In the first stage of the error detection process the checker examined the sample of cards marked by each person, correcting and recording all errors made. Where consistent errors were revealed, the marker (whose initials appeared on the card) was consulted and an attempt made to improve the quality of the marking. An example of the efficacy of this method was seen in the case of one marker who on the third day of the survey had an error rate of 2.7%. The marker's attention was called to the type of error, the mistake was realized, and further checks revealed no error of this type in the marking by this person.

The marking of the book card information on the IBM cards was done under optimum conditions. Lighting was good, each person had his own work, and the only tool the marker used was the book card itself. Under these conditions a spot check was easy to carry out and close supervision made it possible to keep errors at a minimum. The error rate for this section was 1.0%. Borrower and shelf list information had to be marked under adverse conditions. The workroom was crowded, lighting was poor, and it was necessary for the markers to use a variety of tools to locate information. Working in teams expedited the marking but made individual supervision of the marking difficult. As a result of these factors the rate of error was higher in these two sections -- 1.2% for borrowers' and 2.3% for shelf list information. The rate of error in alphabetizing the fiction cards and separating repeat circulations and duplicate copies was 0.4% while a 1.4% error rate was disclosed in the marking of overdues.

The detection of error involved more than the daily checking of the marking of individuals and teams. A considerable refinement of the marking of the cards was accomplished through the detection of errors by the machines themselves as well as by individuals who noted and corrected errors as they handled the cards in the various processes. Periodically, usually at the end of each day, the marked cards were taken to be punched. The checker usually remained during this process so that he could record the types of errors being made and by whom they were being made. These errors were usually mechanical, and if they could be corrected conveniently the checker did so. Cards with significant errors were returned to the markers for correction the next day. In either case, the errors were called to the attention of markers so that further errors might be minimized. While the above errors detected by the machines were chiefly of a mechanical nature, it was possible through a series of internal checks to utilize the machines in the detection of other errors by the markers. When all the cards had been marked in all columns and, just prior to cross-tabulations of the data by the machines, the internal checks devised earlier were utilized in machine operations. An internal check is a matter of verifying one item of information on the IBM card against another column or group of columns containing information compatible with that in the column being checked. For example, if all cards indicating housewife ("23" in columns 28-29) were separated, the operator ought to be able to stick a needle through the entire deck of those cards in that column (30) and hole (1) assigned to borrowers of the female sex. Obviously if any cards marked 0 (male) in column 30 turned up, a mistake
could be assumed and the original source of information would have to be re-examined.

A more obvious type of internal check is that against holes in columns where it is
known that no marks should have been made. For example, the column dealing with re-
serves and renewals utilized only places 0, 1, and 2 so that any card with punches in
holes 3-9 would be in error. Twenty-five such internal checks were used in the pro-
gram of error detection described, ranging from the very simple type of check to the complex.

Exclusive of mechanical errors, a total of 119 errors were so discovered and corrected
before cross-tabulations were made.

The over-all error rate for the marking of the IBM cards in this survey, as re-
vealed by the approximate 10% sample check made by the checker, was 1%, a low enough
figure to assure reliability of the recorded data. This figure does not take into con-
sideration the refinement of the marking as it was accomplished in the corrections made
after the disclosure of errors by means of the internal checks and through the inci-
dental discovery of errors as individuals worked with the cards. It would seem safe to
say that the total rate of errors in the data as now recorded is probably less than 1%,
when all these factors are taken into consideration. As has been stated, errors were
revealed by three means: by having one individual check a sample number of cards, by
the machines while punching the cards and through internal checks, and by various indi-
viduals working with the survey data who encountered and corrected errors incidentally
in the course of their work. No one of these measures is sufficient unto itself and
each must be dependent upon the other two. Of the three, the first is undoubtedly the
most important and provides a systematic means of computing an error rate for the en-
tire project. However, it involves only a sample number of cards, and while it reveals
and corrects errors found in those cards and tends to improve the quality of the mark-
ing, it does not reveal errors in a large percentage of the cards. These errors must
be detected by the machines through routine use of internal checks, and incidentally
by individuals working on the project. Therefore, any plan for the detection of error
should employ a combination of these three methods.

The Use of the Machines

A knowledge of the capabilities of the International Business Machines will re-
veal many possibilities for using them economically and advantageously. For instance,
the date of circulation for one day can be punched automatically on that group of cards
without any mark-sensing. To give another example, if a borrower took out more than
one book only one card need be marked with the information concerning that borrower
(residence, sex, family status, occupation, etc.). The machines will automatically
punch the rest of the cards for books borrowed by him from the information recorded on
the first marked card. Therefore, although 4,094 individual items circulated, only
1,055 (individual borrowers') cards had to be marked with information about borrowers.
When the cards were being marked for borrower information the machines arranged the
cards by borrowers' numbers to aid in using the registration files. However, when mark-
ing cards for book information, arrangement by type of material was essential. The
juvenile non-fiction and adult non-fiction were separated out by the sorter and arranged
according to Dewey classification to speed the checking of the shelf list. And while
working with this section it was often an advantage to have cards arranged by date of
publication in order to complete information on pagination through the Cumulative Book
Index. In all these procedures the use of these machines was essential and saved hours
of labor.

In planning the tabulation of the survey information, it is desirable to make
efficient use of the data accumulated. To this end, a series of questions refined to
specific terms was compiled. These questions were formulated to extract every possible
useful piece of information from the data on the cards. With these specific questions
in mind, the headings for various columns used in the code were listed on separate cards
and under each heading were listed those items of information against which that particular item in the code was to be cross-analyzed.

For example, at the top of one card "Occupation, columns 28-29" was noted, and beneath were listed "type of material, column 1," "sex, column 30," "family status, column 31," "geographic radius, column 21," etc. This directed that for each occupational group the number of circulations be indicated for each available category of the designated characteristics. The cross-tabulations were made by the machines, and the code designation of the information was hand-labeled on printed IBM tabulation sheets. Thus, for example, the tabulation by cross-analysis revealed information concerning both the volume of books borrowed and the number of individual borrowers. Though 1,625 titles (70% of the total adult books circulated) were fiction as against 711 titles (30%) of non-fiction, 460 people (57% of all adult borrowers) took adult fiction books and 351 people (43%) non-fiction.

Of great help throughout this survey, both in its planning and actual conduct, was the advice and technical skill of Leonard Staugas who did all the work of punching and tabulating the cards. He used the equipment in the University of Illinois Tabulating Office, which was made available without charge. When such facilities are available, or when time is short and extensive tabulations are desired of many data, the marksensing method described above is recommended for studies such as this.

The experience of those who worked with the Urbana survey indicate certain recommendations that would improve such a project:

1. The days of circulation selected for testing should fit availability of markers for a convenient work schedule rather than developing a work schedule to fit an arbitrary sequence of days to be marked. This would allow a minimum number of markers and simplify instruction, supervision, and personal error.

2. It is essential that the pre-tests be carefully undertaken to discover all possible inconsistencies in available information in order to develop clear-cut instructions to the markers. The pre-tests are also needed to establish estimates of time necessary for work schedules.

3. Checks for error should be incorporated in the project so that errors may be detected readily and the information recorded at the various stages of the survey will be reliable. This should include continuous cross-checking of the IBM cards, and the appointment of an individual to be in charge of error detection and to serve as liaison between the markers and the IBM operator.

4. Sufficient time should be allotted all phases of the project to insure unhurried and sure marking, careful alphabetizing of cards, and consistent check for error.

5. Work facilities should be arranged so that adequate lighting and space are available. These are essential for an easy flow of work and for low rates of error in marking the IBM cards.

6. Every card should be initialed by the markers at each phase to insure individual check so that no misunderstanding of instructions can continue after serious errors are caught.

7. Complete understanding of the ways in which the International Business Machines can help the project must always be kept in mind in planning the various phases of the survey.

* * *