DOCUMENT RESUME

ED 032 787 EM 007 471

By-Hansen, Duncan N.; Dick, Walter
The Data World of CAI. Tech Memo No. 8.
Florida State Univ., Tallahassee. Computer-Assisted Instruction Center.
Spons Agency-Office of Naval Research, Washington, D.C.
Pub Date 15 Sep 69
Note-20p.
EDRS Price MF-$0.25 HC-$1.10
Descriptors-*Computer Assisted Instruction, Computer Programs, Cybernetics, *Data Analysis, Data Processing, Electromechanical Technology, *Electronic Data Processing, *Information Processing, Input Output Analysis, Programing, Programing Languages, Relevance (Information Retrieval), Research Problems, Research Tools, *Sorting Procedures, Taxonomy

Misunderstandings on the part of the researcher as to the computer's capability, coupled with the inability of the computer system to function efficiently, has often handicapped the research activities in computer assisted instruction (CAI). A new data system is being developed to resolve these problems. A Delete/Sort program has been designed to allow the author to obtain only relevant information, and to specify output and analysis parameters. To maximize the potential of the IBM 1500/1800 for CAI use, an overlapping input-output routine is to be developed. Since CAI response data is already partially ordered, the use of an internal sequence sensitive bi-directional sort procedure may be advantageous. To improve the data management and analyses operation, in order that more sophisticated research may be performed, the role of the researcher will be studied as he executes the primary steps in utilizing the CAI system--(1) preparation of the experimental materials, (2) the monitoring of the experiment itself, and (3) the supervision of the post experimental data analysis. (JY)

CAI CENTER

TECH MEMO

THE DATA WORLD OF CAI

Duncan N. Hansen and Walter Dick

Florida State University

Tech Memo No. 8
September 15, 1969

The work reported herein was performed in the Florida State University Computer-Assisted Instruction Center under Office of Naval Research N00014-68-A-0494. Reproduction in whole or in part is permitted for any purpose of the United States Government. This document has been approved for public release and sale, its distribution is unlimited.

FLORIDA STATE UNIVERSITY

Tech Memo Series

The FSU-CAI Center Tech Memo Series is intended to provide communication to other colleagues and interested professionals who are actively utilizing computers in their research. The rationale for the Tech Memo Series is three-fold. First, pilot studies that show great promise and will eventuate in research reports can be given a quick distribution. Secondly, speeches given at professional meetings can be distributed for broad review and reaction. Third, the Tech Memo Series provides for distribution of pre-publication copies of research and implementation studies that after proper technical review will ultimately be found in professional journals.

In terms of substance, these reports will be concise, descriptive, and exploratory in nature. While cast within a CAI research model, a number of the reports will deal with technical implementation topics related to computers and their language or operating systems. Thus, we here at FSU trust this Tech Memo Series will serve a useful service and communication for other workers in the area of computers and education. Any comments to the authors can be forwarded via the Florida State University CAI Center.

Duncan N. Hansen
Director
CAI Center

ABSTRACT

DATA WORLD OF CAI

The research activities in CAI have often been handicapped because of the inability of the computer system to function efficiently plus misunderstandings on the part of the researcher as to the computer's capability. This paper describes how the FSU-CAI Center is developing a new data system to resolve these problems. This system will have overlapping input-output routines and internal sequence sensitive sort procedures in order to improve the efficiency of managing the response records from a CAI study. In addition, the research designed for studying the requirements of various categories of researchers who use the FSU-CAI System is explicated. This development will require many user-supported programs which are not in existence today. The purpose is to improve the data management and analyses operation in order that more sophisticated research may be performed.

THE DATA WORLD OF CAI

Duncan Hansen and Walter Dick

The prime user of many CAI systems is the researcher/evaluator. It is he who both creates the learning environment for the student and makes demands upon the system to assist him in analyzing his results. While, on the one hand, the researcher/evaluator may be a conceptualizer of experimental materials and data analysis routines, he is at the same time totally dependent on the students and the data which they generate and the operational staff and the programs of analysis which they produce.

This paper will review our approach to response data analyses and management for CAI. Specific plans for 1969-70 will be presented in order to communicate the essence of our interests. After setting a broad context for CAI research and data analyses, the existing Florida State University Data Management System will be reviewed. This will be followed by a discussion of new computer systems development. And finally by a description of our plans for studying and facilitating the role of the researcher in CAI data analysis.

Two current activities at the FSU-CAI Center are prototypic of the type of needs generated by the researcher. Several projects are currently underway to investigate the relationship between anxiety and performance in learning situations. These studies require the presentation of meaningful learning materials to students via the CAI terminal and the continual monitoring of their anxiety state. Researchers associated with these projects are dependent upon the data analysis system to produce summarized records of the students' performance and an indication of their anxiety states. Attempts are underway to manipulate these two variables in order to maximize student learning.

A second example of the needs of the researcher are evident in the intermediate Science Curriculum Study project. This project, housed in the Department of Science Education at Florida State, has as its mission the development of self-paced science materials for grades 7, 8, and 9. While these materials are designed for implementation in conventional classrooms, they are being evaluated, in part, via CAI. The semi-programmed instructional materials have been prepared for the CAI system with additional questions embedded in them. Forty students from the University School received all their science instruction at the CAI Center. In addition to the instruction at the terminal, the student also conducted numerous experimental activities within the laboratory at the Center. The purpose of this study is to derive learning data from a sample of typical Intermediate Science Curriculum Study (ISCS) students in order that revisions may be made in the curriculum materials. Elaborate coding schemes have been developed in order to quickly and easily summarize the learning data in order that content experts may pin-point student learning difficulties.

The studies listed above, namely, the anxiety investigations and curriculum evaluation, are characteristic of many such studies which are being currently conducted within the FSU CAI Center.

The anxiety studies are being conducted by personnel who are primarily associated with the CAI Center, while the ISCS evaluation is directed by the ISCS project staff. The anxiety studies typically have a duration of one to three hours, while the ISCS study extends over each and every school day from September until June. Though each study has its own unique characteristics, the researchers who are involved have a common set of expectations in terms of "what the system will do for me." Many of the initial expectations have thus far proven to be much more difficult to achieve than was at first anticipated.

The early literature in computer-assisted instruction held great promise for the researcher in terms of the ease with which he could prepare and run an experiment and analyze his data. One could almost assume that the researcher would merely enter his instructional materials into the system and that somehow he would magically be presented with results of this study neatly printed out and ready for publication. This certainly has not proven to be the case. Rather than try to determine the locus for the blame for the lack of data from CAI, it would seem more profitable to describe the activities which Florida State has explored and proposed to explore in order to derive more information about the researcher's use of a CAI system.

Turning now to the FSU Data Management System, the response record recorded by the 1500 system is a variable length record containing such items as the response itself, the latency, the contents and status of various record-keeping storage areas and an identifier for the response record. In order to maximize efficiency, these variable length records are first converted by the FSU Data Management System into condensed standard length records. Where necessary, continuation records are generated. These records are then sorted according to the course, the student, the record type (that is, whether it is a course header, a student header, or a response record), the date, the time of day, and whether it is a continuation record. The sorted records are then merged into a master file that constitutes the basic file with which the remainder of the programs work.

Next in the development of the Data Management System was the writing of what is called the Delete/Select program. In many ways, the uniqueness and power of this system resides in this program. It is through the use of the Delect/Select program that the author is able to obtain only that information which is relevent to answering questions regarding the data and not be overwhelmed with an unmanageable mass of data. It is also possible to specify output and analysis parameters so that the data is then available either as printed output in summarized records ready for the author to use, or as tape or punched records which then will be considered raw data for statistical analysis routines. The set of programs to provide output in the form of raw data for use by statistical programs are now available. We turn now to our plans to improve this system.

CAI Operating Programs and Systems

Computer Science which includes the applied aspects in current hardware and language systems represents a generic conflict between abstract models that stress generality as opposed to restricted empirical applications that stress specificity for efficiency's sake.

CAI systems are an example of an empirical application requiring effective operations. It has been observed that most CAI response data has gone unanalyzed due to the unavailability of both computing time, (CAI systems tend to have small central processing units [CPU]), and systems programs that allow for specific types of data manipulation. We propose to study two factors within CAI operating systems in order to assess the need for highly efficient and specific programming, namely, improved input/output data transfers and a sort specially designed for CAI.

Input-Output Routines

In the FSU Data Analysis and Management System in which large files are regularly manipulated, highly efficient programming techniques have been utilized to maximize program execution times. Programs in this system can normally execute no faster than the time it takes to pass data through it; hence, the term "throughput" is derived. The minimum execution time for such programs can be no less than the maximum speed at which the slowest limiting input/output (I/O) devices can pass the data. Throughput improvements are aimed at speeding up internal execution speeds in order that a program will approach I/O boundary conditions while maximizing efficiency. It is the improvement in this overall computer throughput that will allow CAI systems to provide significant capacity increases and facilitate greater operating economies.

Under the proposed development for 1969-70, I/O operations on the 1800 will be designed to maximize the hardware's potential capabilities. It is our contention that IBM did not provide for this efficiency when developing their 1500/1800 systems especially in background applications. The IBM 1800's I/O channels operations could be fully overlapped, simultaneous activities. To implement this, multiple buffering spaces for I/O must be employed in order to allow CPU operation on a data block while data channels are transfering between other I/O buffers. An additional ingredient will be the software control of I/O space utilization by queuing techniques which greatly simplify the ordering decisions process. The result should be a total operating system that always has data on which to work. Thus, improved overlapping I/O routines for the 1500 CAI system could well be a required design characteristic for all future CAI systems.

An Internal Sequence Sensitive Bi-Direction Sort

The development of this sort is proposed for development in order to take advantage of the fact that CAI data records are not randomly stored but rather are more often only slightly out of sequence. The bi-directional sort should locate the correct sequence position of a currently encountered out-of-sequence record in no more than eight comparisons for any 100 response records. Intermediate data swaps will be eliminated by indirect data examinations via a bi-directionally linked index listing procedure.

For clarity's sake, a comparative description of the present IBM sort will be presented. Our current sort technique can be described as a selection and interchange method (Douglas, 1959). Imagine an unordered string of data items presented with reference to a sorting pointer. All data to the left of the pointer will have been ordered, while that to the right (including the pointer) remains to be sorted. Commencing from the left with the pointer at the first item, all data to the right of the pointer is serially compared to and interchanged with the pointer item, if necessary. After all data has been examined (a scan completed), the pointer is advanced to the second item and another scan pass is made. When the pointer finally points to the last item in the data string, the internal sort is completed. It is clear, therefore, that the feature of partial ordered data as in CAI response records will be ignored. The sort will require

$$\sum_{i=1}^{n} i = \frac{n\,(n-1)}{2}$$

comparisons to be complete. Also, the swapping required to move a data record during the sorting process may involve several computer words. This is time consuming and inefficient. Since a record's position in the sorted string of records is not fixed until the final sort pass is completed, a record most likely will change position several times. These intermediate record swaps are both premature and costly, especially given the partial ordering of CAI response records.

The proposed internal sequence sensitive bi-directional sort will utilize and integrate three effective and complementing techniques: (1) repeated comparison and interchange of data (see Shapin, 1954 and Bell, 1958), (2) Dedekind cuts, or "binary chopping" (Hardy, 1945), and (3) bi-directionally linked indirect index lists of data manipulation.

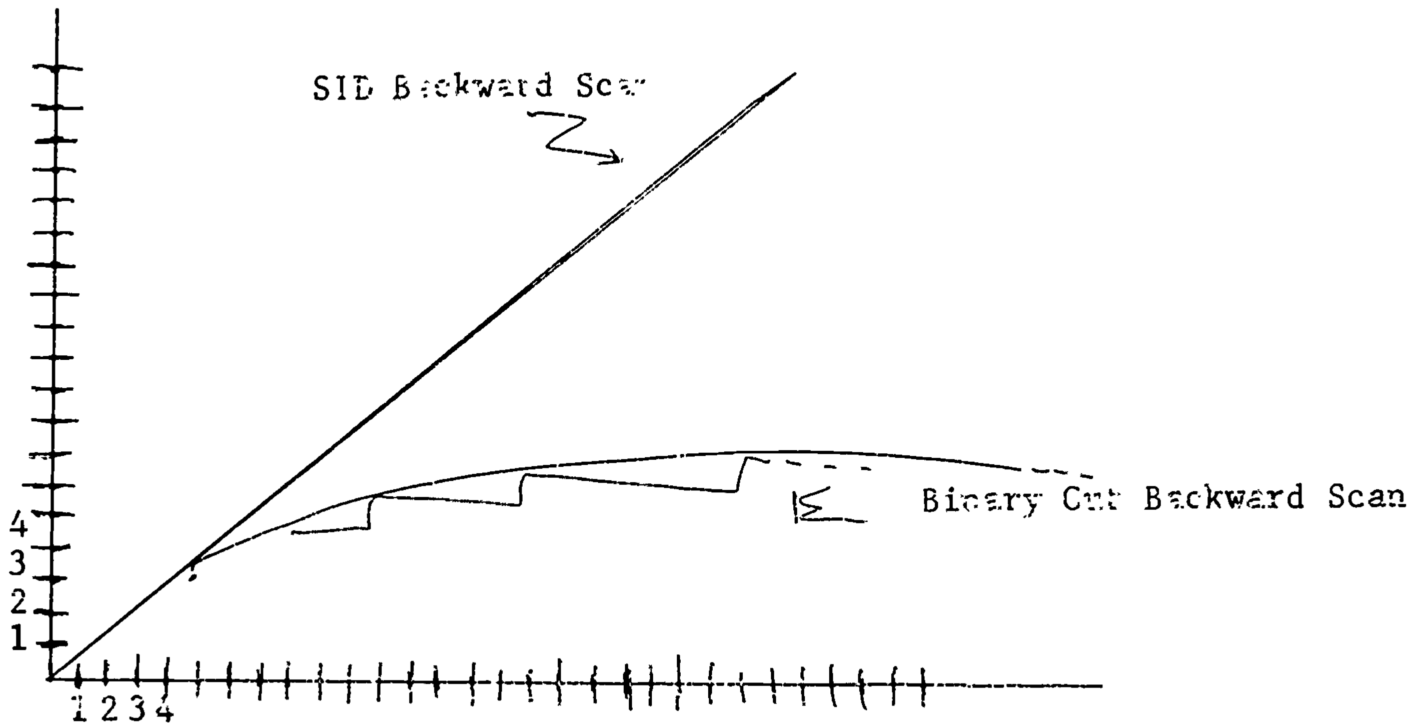
The repeated comparison and interchange technique utilizes the sequences that already exist in a data block to reduce the number of comparisons necessary to order the string. Utilizing a pointer, a comparison is made with each subsequent item following the pointer. If the sequence is proper, the pointer is stepped up by one and the comparison operation is repeated. If the sequence is unordered, the present position of the pointer is remembered, and the data items are interchanged. The pointer positioning is then decremented by one (but cannot be positioned less than at the first item of the string) and another comparison operation is performed. Once the position of the out of sequence record is established on the backward scan, the forward scan re-commences from the remembered position of the pointer It is important to note that data to the left of the pointer is ordered and to the right is unordered.

At this point the observation can be made that with respect to number of required comparisons the best case of the conventional method is equal to the worst case of the proposed method. If we let x = number of comparisons actually required, n = the number of records, and k = the number of initial good sequences (strings) such that $1 \leq k \leq n$ then, $\lim_{k \to 1} X = n-1$

Thus this comparison technique insures equivalent efficiency to the existing 1500 sort.

Dedekind cuts, or "binary chopping" is derived from the observation that when J equals the least integer such that $2^J \geq n$ where n is the number of elements in a list, then every element in that list can be examined by making no more than J halvings of n. Using that value of J to proceed in the direction of locating a targeted item and given that the items to the left of the pointer are ordered, it is clear that the proper direction can be established between an out-of-sequence data item and its potential sorted position in the list. Utilizing this fact, we can be assured of locating the proper sequential position for a data item that is out of sequence within L tries where P = current position of pointer and L = least number such that $2^L \geq P$. For example, if the 23rd record is being compared to the 24th, the position of the 24th record can be determined in five comparisons as opposed to 23 in the conventional sort. Figure 2 offers graphic representation of the efficiency of the conventional and the Dedekind cut techniques. Thus this Dedekind cut search technique will facilitate the efficiency of the sort.

Figure 2



A bi-directionally indirect linked index listing procedure is desirable for two reasons. First, sorting can be achieved by simply finding a record's logical position in the list and then linking the item into position without ever actually moving data, which is a time-consuming routine. Second, due to the linking procedure, the indexes themselves do not move, only the sort pointers change. At the completion of the sort, it remains only to find the first logical element in the index list, and build an output array of data by scanning forward through the index list while transferring the data referred to by the index to the output area. Even this inconvenience would not be necessary if the 1800 computer was capable of scatter writing to disk instead of only to magnetic tapes. Thus, the bi-directional linked index list increases efficiency by moving data only at the completion of the sort procedure.

The development and availability of the overlapping I/O routines and the bi-directional sort will provide an opportunity to evaluate the extent to which CAI systems should have specific programs that capitalize on the unique features of CAI operations, namely, high I/O requirements and partial ordering of CAI response data. By comparison with our existing 1500 system with its sort, we can empirically measure the improved operational efficiency due to each development as well as the combined effectiveness of all new developments. Thus we will have an empirical basis upon which to make the potential recommendation that existing or future CAI systems employ these two types of systems features.

Proposed Research Activities

A preceding section described the activities of the anxiety researcher and the curriculum development researcher. There are, in fact, many varieties of researchers who find themselves wanting to use the CAI system as a research tool. In utilizing the CAI system, the researcher must be concerned with three primary steps: (1) the preparation of the experimental materials, (2) the monitoring of the experiment itself, and (3) the supervision of the post-experimental data analysis. Each of these steps is crucial to the successful completion of the experimental research, and therefore, each must be carefully designed and executed. It is the purpose of our future FSU activities to examine the role of the researcher at each one of the steps in the preparation and analysis of a CAI research study. It is hypothesized that the derivation of demographic data will provide other researchers with a more enlightened view of CAI research and a more realistic expectation of the competencies which he himself must have. A direct outcome will be a document which will outline recommended training procedures for future CAI researchers.

The first hurdle which the potential CAI researcher faces is that of learning a CAI language. It may be argued that the researcher need have no technical competency in the instructional language, but rather, he should conceptualize his instructional problem and have a professional coder prepare the materials for the CAI system. Or, it may be argued that in order to fully utilize the capabilities of the CAI system and to likewise recognize its limitations, a researcher should have some knowledge of the CAI language which is to be employed. Too often, the end result is that the researcher either spends a great deal of time learning the entire CAI language when he needs to know only parts of it, or he learns none of the CAI language and consequently, his communication with professional coders is greatly limited and the time span required in order to implement his materials is greatly lengthened.

A seemingly reasonable alternative procedure would be to have the researcher learn only those capabilities of the CAI language which will be of most value to him in his research. However, at the present time there is no data which would indicate what parts of any CAI language are most needed and used by the researcher. It is, therefore, proposed that a taxonomy be established which would document the relative frequency of CAI language characteristics which are most often utilized by researchers. More explicitly, this taxonomy will classify various research efforts into similar categories and then indicate the relative frequency of the researchers' use of such language capabilities as: (1) implicit branches, (2) explicit branches, (3) labeling, (4) use of functions, (5) random number generator, (6) counters, (7) switches, (8) graphics, (9) dictionary characters, (10) macros, and (11) response analysis techniques. The primary purpose of this taxonomy will be to determine if there are characteristics of the CAI language which are more likely to be used by one type of researcher than another. The ultimate outcome will be the streamlining of instruction for the potential CAI researcher and the facilitation of his communication with the professional CAI coder.

It should be noted that after the coder enters the instructional materials on the CAI system the researcher is not yet ready to carry out his experiment. He must first, in some way, check to make sure that the program is functioning exactly as he has designed it. After the experiment has begun, the researcher should avoid editing the instructional materials. Therefore, it is necessary to determine those techniques which may be most useful in determining when the program is ready for actual student use. We plan for a special "debugging" program to be developed which will indicate to the author the following: (1) the frame or labeling identifiers he has used, (2) the enter and process identifiers and the answer identifiers he has used, (3) those points in the instructional program in which he has used counters, and (4) those points in the instructional program in which he has used switches. An attempt will also be made to create a "dummy student" program. This program would in essence execute the author's program as if he were a particular type of student. That is, the "dummy student" could get every answer cor-rect, or every answer incorrect, or execute all statements which set switches, etc. The researcher could then check the various switch and counter settings in order to determine if the program had executed correctly.

To summarize the first portion of the research related to the preparation of the experimental materials, the taxonomy of research activities will be derived which indicates the relative frequency with which the various CAI language characteristics are utilized in the various types of research projects. The second aspect of this area will be the development of various course listings and "dummy student" programs which will facilitate the researcher's effort to have his instructional materials as nearly error free as is humanly possible.

The second major concern of the researcher is the actual monitoring of the experimental activities. The techniques which one might wish to use on order to monitor the proceedings might differ depending upon the actual length of the experimental situation. For example, considerably different procedures would be used for a one hour, one-shot experimental study as opposed to a year-long instructional activity.

In the former situation, as a one hour study, the monitoring becomes quite crucial. In this case, a check must be made during the actual instruction or immediately following it in order to determine if the session has gone as planned. In order to implement on-line checking of student activities it is proposed that a station-to-station function be implemented which would permit the researcher to sit at a particular terminal and monitor the activities which were going on at the various learning stations. In addition, he will be provided the opportunity, immediately at the end of the experimental session, to check the settings of his subjects' counters and switches in order to determine that the proper number of responses has been made. Those studies which require many hours of learning from the student can utilize some of the data management techniques which are documented in the next section. These procedures can be utilized to identify students who are not readily adapting to the learning setting.

Therefore, in part two of the research it is proposed that several within experiment and immediately following experiment functions be implemented in order to provide the researcher with immediate confirmation that the experimental materials have been executed properly. These and other techniques utilized by researchers in order to maximize the reliability of their results will be documented, and the relative frequency of the use of each technique will be indicated. Unsuccessful monitoring techniques will also be documented with indication of why they were unsuccessful in detecting improper procedures.

It is also proposed that the researcher participate in the data reduction and analysis phase of this project. The first would be as a colleague in the development of the extensions of the FSU CAI Data Analyses and Management System. As outlined elsewhere in this paper, this will include the development of an edit routine which will further refine the data for the researcher. As a result of this program, various print-out summaries will be required as well as punched cards containing summarized data. It is anticipated that the data punched on cards will be submitted to conventional statistical analysis routines which are already available either on the CAI Center's IBM 1800 system, or on FSU's Control Data 6400 system. In addition, all research users of the CAI system will be asked to document their complete needs for data analysis and to indicate those programs which they feel will be most useful for their purposes. If at all possible, those routines which seem most useful will be developed. In addition, newer analysis techniques such as sequential testing procedures, which may be of benefit to users who are unaware of their availability, will be implemented. Thus, all users of the data analysis routines in the CAI Center will be asked to contribute either directly or indirectly to the future development of data analysis routines.

Secondly, it is proposed that frequency counts be made of the use of each of the various analysis programs which are developed. At the present time there is no known data on the frequency with which various types of analysis routines are utilized by researchers. This frequency count should be invaluable in terms of establishing priorities for programming needs for other CAI installations in the future.

As a logical conclusion to the development of programs and the counting of the frequency with which they are used, subjective reports will be requested from each researcher after he has received data from the CAI Data Management System. These reports will reveal the usefulness of the data to the researcher, and request suggestions about ways in which the programs can be enhanced or extended. This information will be continually summarized and reported back to the programming staff in order that the entire programming development effort may be characterized as a re-cycling and continuously updating process as opposed to an open-ended procedure.

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Florida State University Computer-Assisted Instruction Center Tallahassee, Florida 32306	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

THE DATA WORLD OF CAI

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

TECH MEMO NO. 8, SEPTEMBER 15, 1969

5. AUTHOR(S) (First name, middle initial, last name)

DUNCAN N. HANSEN AND WALTER DICK

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
September 15, 1969	15	0

8a. CONTRACT OR GRANT NO. N00014-68-A-0494 b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY OFFICE OF NAVAL RESEARCH MAIN NAVY BUILDING, WASHINGTON, D.C.

13. ABSTRACT

The research activities in CAI have often been handicapped because of the inability of the computer system to function efficiently plus misunderstandings on the part of the research as to the computer's capability. This paper describes how the FSU-CAI Center is developing a new data system to resolve these problems. This system will have overlapping input-output routines and internal sequence sensitive sort procedures in order to improve the efficiency of managing the response records from a CAI study. In addition, the research designed for studying the requirements of various categories of researchers who use the FSU-CAI System is explicated. This development will require many user-supported programs which are not in existence today. The purpose is to improve the data management and analyses operation in order that more sophisticated research may be performed.

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT

DD FORM 1 NOV 65 1473 (BACK)

S/N 0101-807-6821

Security Classification

A-31409