Order Number 9014627

A comparison of manual/electronic versus electronic approaches to the development of information retrieval skills in electronic database instruction

Dudley, Suzanne Louise, Ph.D.

University of South Florida, 1989

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

A COMPARISON OF MANUAL/ELECTRONIC VERSUS ELECTRONIC APPROACHES TO THE DEVELOPMENT OF INFORMATION RETRIEVAL SKILLS IN ELECTRONIC DATABASE INSTRUCTION

by

Suzanne L. Dudley

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Curriculum and Instruction in Interdisciplinary Studies in the University of South Florida

December 1989

Major Professor: Andria P. Troutman, Ed.D.

Graduate Council
University of South Florida
Tampa, Florida

CERTIFICATE OF APPROVAL

Ph.D. Dissertation

This is to certify that the Ph.D. Dissertation of

Suzanne L. Dudley

with a major in Curriculum and Instruction has been approved by the Examining Committee on November 27, 1989 as satisfactory for the dissertation requirement for the Ph.D. degree.

Examining Committee: Andria P. Troutman

Major Professor: Andria P. Troutman, Ed.D.

James C. Dickinson

Member: James C. Dickinson, Ph.D.

E V Johanningmeier

Member: Erwin V. Johanningmeier, Ph.D.

Constance V Hines

Member: Constance V. Hines, Ph.D.

J M Knego

Member: John M. Knego, Ph.D.

ACKNOWLEDGEMENTS

I have greatly appreciated the support and encouragement that I have received from faculty and family during my doctoral program. I would like to thank the members of my program committee and my dissertation supervisory committee for their interest and guidance.

My family, Alan, Rachel, Amanda and Elizabeth, deserve special recognition and grateful thanks. Without their love, understanding and commitment achieving this goal would have been almost impossible. They are very special people!

TABLE OF CONTENTS

LIST OF TABLES iv

LIST OF FIGURES vi

ABSTRACT vii

CHAPTER I: INTRODUCTION 1
Statement of the problem 6
Purpose of the Study 7
Research Questions 12
Definition of Terms 12

CHAPTER II: REVIEW OF THE LITERATURE 15
Trends in the Use of Database Technology in Society 16
The Integration of Database Technology into the Curriculum 19
Research into Database Skills 21
The Acquisition of Skills 26
Problem Solving 35
Implications for the Design of Instructional Materials and Treatments 52

CHAPTER III: METHOD 56
Instructional Design 59
Stage One: Instruction for Entry Level Skills 59
Stage Two: Instruction for the Development of Information Retrieval Strategies 63
Stage Three: Instruction for Evaluating Accuracy of Information Retrieved From a Database 72
Statistical Design 74
Participants 79
Data Collection During Stage One Activities 81
Data Collection During Stage Two Activities 81
Data Collection During Stage Three Activities 83
Post Hoc Analyses 84
Statistical Analysis Software and Hardware 85

CHAPTER IV: RESULTS 86
Stage One: Performance on the CRMT 87
Stage Two: Effectiveness of Information Retrieval During Event 1 Introductory Activities 88
Stage Two: Effectiveness and Efficiency of Information Retrieval 91
Stage Three: Effectiveness of Information Evaluation 96
Post Hoc Data Analysis 97

CHAPTER V: SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS 108
Conclusions 111
Implications 117
Limitations of the Study 119
Suggestions For Further Research 120

REFERENCES 124

APPENDIXES 137

A. Instructional Materials: Database Concepts and Skills 138

B. Instructional Goal, Instructional Analysis, and Performance Objectives 173

C. Criterion Referenced Posttest for Instructional Materials 180

D. Data Set 188

E. Specifications for Information Requests in Sets A - D 195

F. Information Requests for Sets A - D 196

G. Sample Strategy Guide 198

H. Sample Menu Screens 200

I. Coding Criteria for Sets A - C 203

J. Sample Evaluation Checklist 209

K. Listing of dBase III Plus Command File User-Interface 211

LIST OF TABLES

1. Summary of the Sequence, Type of Retrieval Activity and Database Medium for Instructional Events in Stage Two 70

2. Descriptive Statistics of Participant CRMT Scores by Mastery Level and Group 87

3. Summary Statistics for Process Retrieval Effectiveness Ratio Scores During Stage Two Introductory Activities 89

4. Summary Statistics for Product Retrieval Effectiveness Ratio Scores During Stage Two Introductory Activities 90

5. Analysis of Variance of Product Retrieval Effectiveness Ratio Scores by Group During Stage Two Introductory Activities 91

6. Summary Statistics for Process Retrieval Effectiveness Ratio Scores During Stage Two Assessment Activities 92

7. Analysis of Variance of Process Effectiveness Scores by Group With Repeated Measures on Set 93

8. Summary Statistics for Product Retrieval Effectiveness Ratio Scores During Stage Two Assessment Activities 94

9. Summary Statistics for Process Information Retrieval Efficiency Scores During Stage Two Assessment Activities 95

10. Summary Statistics for Product Information Retrieval Efficiency Scores During Stage Two Assessment Activities 96

11. Summary Statistics for Evaluation Effectiveness Scores During Stage Three Activities 97

12. Distribution of Participants Across Treatment Groups and Mastery Level 99

13. Means and Standard Deviations of Information Retrieval Effectiveness Scores Across Mastery Levels and Treatment Groups 99

14. Means and Standard Deviations of Information Retrieval Efficiency Scores Across Mastery Levels and Treatment Groups 100

15. Means and Standard Deviations of Information Evaluation Effectiveness Scores Across Mastery Levels and Treatment Groups 100

16. Analysis of Variance of Process Effectiveness Ratio Scores by Mastery Level During Stage Two Introductory Activities 101

17. Analysis of Variance of Process Effectiveness Ratio Scores by Mastery Level During Stage Two Assessment Activities 102

18. Analysis of Variance of Process Efficiency Ratio Scores by Mastery Level During Stage Two Assessment Activities 102

19. First Order Partial Correlations Between CRMT Scores and Subsequent Information Retrieval and Evaluation Performances 103

20. Mean Group Process Retrieval Effectiveness Scores For Each Information Request During Stage Two Assessment Activities 104

21. Mean Accuracy Ratios for Performance on Information Retrieval Component Tasks 105

22. Analysis of Variance of Sort Effectiveness Ratio Scores by Mastery Level During Stage Two Assessment Activities 106

23. Analysis of Variance of Selection Effectiveness Ratio Scores by Mastery Level During Stage Two Assessment Activities 106

24. Analysis of Variance of Field Specification Effectiveness Ratio Scores by Mastery Level During Stage Two Assessment Activities 107

LIST OF FIGURES

1. Design of the Three-Stage Instructional Sequence 58

A COMPARISON OF MANUAL/ELECTRONIC VERSUS ELECTRONIC APPROACHES TO THE DEVELOPMENT OF INFORMATION RETRIEVAL SKILLS IN ELECTRONIC DATABASE INSTRUCTION

by

Suzanne L. Dudley

An Abstract

Of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Curriculum and Instruction in Interdisciplinary Studies in the University of South Florida

December 1989

Major Professor: Andria P. Troutman, Ed.D.

This study examined a manual/electronic versus an electronic only approach to training novice database users in information retrieval from an electronic database. Eighty-two undergraduate university students were randomly assigned to one of two treatment groups: manual then electronic (MTE) or electronic only (EO). All participants completed instructional activities during Stage One of the study using a systematically designed instructional package which presented skills and concepts that had been defined as prerequisites for information retrieval. Participants then completed three sets of instructional activities during Stage Two of the study in which they developed, consolidated and applied information retrieval strategies using an algo-heuristic model of information retrieval. During these instructional activities measures of retrieval effectiveness and retrieval efficiency were taken. Stage Three of the study required participants to evaluate the accuracy of retrieved information, and measures of evaluation effectiveness were taken. Although mastery participants retrieving information from a printed database were significantly more effective in generating the required information product than mastery participants retrieving the same information from an electronic database, there was no significant difference between the two mastery groups for measures of process effectiveness, process efficiency or evaluation effectiveness. These results indicate that there is no advantage, or disadvantage, in including manual information retrieval activities when participants are first developing information retrieval strategies.

The statistically significant positive correlations between performance on the criterion-referenced mastery test and measures of information retrieval and evaluation performance provided support for the requirement for mastery of the defined entry level skills and concepts before participants begin electronic information retrieval. The study findings indicate that by achieving mastery of the defined entry level skills, and applying these skills using an algo-heuristic model for developing information retrieval strategies, some novice database users can achieve levels of competence equivalent to experienced database users with just six hours of instruction.

Suggestions for further research address the role of feedback during information retrieval, the use of strategy guides, clarification of the perceived nature of the task to novice computer users, and the application of an algo-heuristic model in more varied information retrieval environments. (148 references)

Abstract approved: ______________________

Major Professor: Andria P. Troutman, Ed.D.
Professor
Department of Instructional Computing

November 27, 1989
Date of Approval

CHAPTER I

INTRODUCTION

The electronic storage of data in database systems, and the retrieval of information from these systems, are integral processes in a wide range of activities associated with research, business and commerce. Katzeff (1986) claimed that

> as the computer revolution spreads its tentacles in modern society, there is hardly one office or institution left that does not make use of at least one type of database system. The nation's social security register, offices' records of employees and their salaries, and banks' records of accounts are but a few examples of database systems encountered in the modern world. (p. 1)

Data may be stored in a range of forms from coded data in numeric format--such as the majority of data collected from census forms--to full text data--such as the text of newspapers, encyclopedias and legal case reports. However, data in a database are only of use when an appropriate subset of that data can be selected, retrieved and organized to provide the user with information specific to the user's current problem (Shoderbek, Shoderbek & Kefalis, 1985). For this reason, terms such as "information retrieval" and "information processing" are used to describe the decisions and processes that are involved in transforming data into useful information.

Dramatic increases in the processing capabilities of microcomputers over the last eight years, combined with the decreasing costs for hardware and the increasing sophistication of software, have provided the opportunity for data storage and information retrieval to become routine classroom activities. These activities have the potential to enhance the acquisition of knowledge within disciplines and facilitate the development of skills for analyzing, classifying, comparing and contrasting, evaluating, generalizing, inferring and extrapolating.

The potential instructional benefits of using database technology were described by Hannah (1987) who claimed that although

> it is tough finding ways to get students to probe into content, ask critical questions, and test their conclusions ... database activities have the potential to aid students in developing these higher-level thinking skills. They can provide opportunities for students to analyze data, develop theories and evaluate their ideas (p. 16).

Similarly, Rooze (1987) stated

> A wealth of opportunity for developing analysis, synthesis, and evaluation skills lies in preparing and searching a database, but teaching the use of databases does not ensure the learning of these thinking skills. If we are teaching thinking, we must focus the student's attention away from the media to the thinking processes we are using, making children aware of the thinking skills that are embedded in our teaching strategies. The teaching strategies we are using need to be translated into thinking strategies. For the computer to become a tool for thinking, the children must be still taught and they must still learn the

thinking skills involved (p. 7).

To focus on the thinking strategies described by Rooze, prior acquisition of electronic information retrieval skills is necessary if searching a database is part of the thinking skills activities. According to Sternberg (1986), such acquisition frees mental resources for the main problem task.

Despite the potential pedagogical value of electronic information retrieval, Collis (1988) stated that "enthusiasm in endorsing the use of data base programs is somewhat tempered by the difficulties we have noticed adults experience when learning to use a database" (p. 51). Hedberg and Perry (1984, p. 18) observed that "novice computer users find interrogation of on-line information systems difficult and approach such tasks with a high level of anxiety ... [and that] most computer naive people employ only the simplest functions of a database query system despite training on the available range of functions". Tapsfield (1985) suggested that many "teachers and teacher training students probably lack the basic information skills that everyone says they should be teaching to their students" (cited in Collis, 1988).

Experiences among educators (Collis, 1988; Hannah, 1987; Hunter, 1984; Lockard, Abrams & Many, 1987; Rooze, 1987) have led to the observation that information retrieval from an electronic database is a problem solving activity requiring the application of a range of cognitive skills. However, many authors assume that teachers already possess these skills and that they are easily able to transmit these skills to students in the schools. As a result, the research literature does not address questions such as, "Why do many teachers experience difficulty in retrieving information from electronic databases?" or "What are effective methods of teaching information retrieval skills?".

In the last decade, there were promises that electronic data retrieval systems would be created having the built-in intelligence of the "expert" in a field (Waterman, 1986). To query such systems, the user would use natural language. Although some specialized expert systems have been developed that rely on natural language processing, the development of this type of application has been impeded by several factors:

1. difficulty in extracting and coding the knowledge of experts,
2. cost of development,
3. the need for sophisticated hardware systems, and
4. inability to achieve reliability (Frenzel, 1987).

Nevertheless, the promise of expert systems that rely on natural language processing may have led many researchers to underestimate the need to conduct research related to the computer user's skills to formulate machine-processible statements. These formal statements are required by the retrieval systems that are currently most widely available, frequently used, and least expensive. These information retrieval systems are classified by Radecki (1983) as data or discrete retrieval systems and are represented by software packages such as Appleworks, PFS File, PC-File, and dBase III Plus.

Currently, conflicting conditions prevail. Electronic computing devices along with discrete information retrieval systems--databases--are becoming widely accepted as tools for learning and teaching. Software vendors, professional educators, and textbook writers claim that teachers in many disciplines should construct problem solving opportunities for their students through the use of discrete information retrieval systems (Troutman & White, 1988; Hunter, 1984). County and state school systems are mandating minimum competency criteria which include the need for students to acquire electronic information processing skills. Yet a common observation made by software trainers is that novice computer users encounter considerable difficulty learning to retrieve information effectively and often face such tasks with high levels of anxiety. These observations, as indicated earlier, are compatible with reports of other researchers (Collis, 1988; Hedberg & Perry, 1984; Tapsfield, 1985). Herein lies a dilemma. How can teachers, who are novice computer users, use available electronic information retrieval resources to design problem-solving activities emphasizing high levels of cognition, when they themselves cannot effectively retrieve electronically stored information and even fear trying? Obviously, they can't use the available electronic resources without instruction and practice. Therefore, another more fundamental question must be answered. How can teachers be efficiently trained so that they can reliably retrieve valid information that has been stored electronically?

Clearly, some light must be shed on this second question before we can expect teachers to use computers to provide challenging learning experiences for their students. Thus, the second question will be the focus of the study.

Statement of the Problem

Understanding the basic dilemma faced by educators who try to use, in the classroom, the same approaches to information processing that are in general use in the wider community, leads the researcher to two questions that need to be addressed. These are:

1. What concepts and skills are needed to gain proficiency in electronic information retrieval?

2. How can novice computer users be efficiently trained to use these concepts and skills to accurately retrieve valid information that has been stored

electronically?

Purpose of the Study

In a report by the National Academy of Education, called Improving Education: Perspectives on Educational Research, Resnick (1984) posed the following three questions that provide "a framework within which fruitful cognitive research on education can proceed". These were (a) what are the mental processes typical of experts, (b) how are processing capabilities of experts acquired, and (c) what are the effects on mental processes of different kinds of instruction?

These questions can provide the framework for translating the problem statement into appropriate statements of purpose for scientific inquiry. To address the conditions of the problem posed in this study, Resnick's questions can be modified to describe three component purposes of this study:

1. To determine what knowledge and skills are required to be proficient in information retrieval from an electronic database,

2. To determine ways in which novice database users can become proficient in information retrieval, and

3. To determine whether specific instructional techniques will increase the information retrieval knowledge

and skills of teachers who are novice database users.

The first component concerns the specification of the knowledge and skills possessed by experts in any given domain. By definition, an expert is one who understands, acts and learns from results without any conscious awareness of the process (Dreyfus & Dreyfus, 1984, 1986). Waterman's (1986) descriptions of the difficulties faced by knowledge engineers in capturing the knowledge of domain experts in expert systems apply to documenting the skills of expert database users. For this study, the researcher analyzed the basic information retrieval knowledge and skills used by experienced database users. This analysis formed the foundation for the systematic design of an instructional package to introduce database concepts, rules, and procedures. Complete details regarding the development, trialing and delimitations of this instructional package are included in Chapter III.

Component two addresses the development of expertise, which Dreyfus and Dreyfus (1984, 1986) view as a five stage process along the continuum from novice to expert. As the development of experts requires extended periods of practice to increase the range and complexity of their performance, this study is limited to the stages: novice, advanced beginner, and competence. The Dreyfus model provides implicit guidelines for structuring content and process activities in the instructional treatments, and in their assessment. For example, the model supports the sequence of (a) presenting context-free rules or facts to novices, followed by (b) presenting real situations requiring application of individual facts or rules to advanced beginners, then (c) presenting more complex situations which require a hierarchical view of decision-making at the third level, competence. The sequence described delineates the types of skills expected at each level of proficiency in skill acquisition.

Component three addresses the potential effect of different instructional treatments on learning outcomes. Varying the treatment received by participants when preparing them for a task such as information retrieval from an electronic database cannot logically involve any variation in the knowledge and skills presented if all are considered to be prerequisites for the desired learning outcome. However, differences in treatment can occur when participants are provided with different media for the presentation of the learning activities. For example, a printed database file provides a more tangible environment in which information retrieval skills can be developed and the effects of decision making during information retrieval can be observed. This is in contrast to an electronic database file which provides an abstract environment where the data may be hidden and where the user is provided with no intermediate feedback on the effects of decision making.

Manipulating a printed database file could provide learners with a concrete analogy of the structure and function of an electronic database that facilitates the formation of the concepts and rules associated with information retrieval. Results of research concerning conceptual models and mental models (Carroll & Mack, 1985; Borgman, 1986; Jagodzinski, 1983; Kahney, 1983; Kahney & Eisenstadt, 1982; Norman, 1983) suggest that the participants whose learning environment provides an understandable conceptual model are able to develop their own dynamic mental models of general world knowledge and domain specific knowledge. In the context of this study, when participants manually retrieve information from a printed database, they experience a tangible conceptual model from which they may then form their mental model. If manipulation of data in a printed database assists participants in developing a mental model and if this mental model incorporates use of previously learned knowledge about information retrieval concepts, then it seems reasonable to expect that gains in the effectiveness and efficiency of the subject's electronic retrieval performance using a computerized database could be observed.

For the purposes of this study, effectiveness of information retrieval is measured by the accuracy with which decisions are made that lead to the retrieval of required information, whereas the efficiency of information retrieval is measured by both the accuracy with which decisions are made leading to information retrieval and the time required to retrieve the information.

Although effectiveness and efficiency are two important measures of information retrieval performance, the novice database users' evaluation of the results of information retrieval is another measure that provides useful insight into their skills. While it is possible to issue a series of commands that are syntactically correct resulting in the retrieval of a set of data, critical skills in evaluating information retrieval results enable learners to identify errors in decision-making and eliminate them in their subsequent retrieval activities.

The three measures of information retrieval performance allow redefinition of the problem statement into four separate research questions for this study. The groups referred to in the research questions are classified in the following way. Participants were assigned to use either a printed or an electronic database during the first stage of this study. The group that used the printed database file for manual information retrieval prior to working with the computerized database for electronic information retrieval was referred to during this study as the manual then electronic (MTE) group. The group that worked with only the computerized database file completing electronic information retrieval activities was referred to as the electronic only (EO) group.

Research Questions

The research questions examined in this study were:

1. Is there a significant difference between measures of manual retrieval effectiveness and measures of electronic retrieval effectiveness during introductory learning activities in retrieval strategy implementation?

2. Is there a significant difference between performance of the MTE group and the EO group on measures of retrieval effectiveness during subsequent information retrieval learning activities in a computerized environment?

3. Is there a significant difference between the MTE group and the EO group on measures of retrieval efficiency during subsequent information retrieval learning activities in a computerized environment?

4. Is there a significant difference between the MTE group and the EO group on measures of information evaluation?

Definition of Terms

To clarify the meaning of terms included in the problem statement, the purpose of the study, and related research strategies the following terms are defined:

A <u>computerized database file</u> is a collection of records consisting of data that were recorded using computer software that facilitates the entering, editing, deleting, ordering and retrieving of data/information.

Information is the subset of data from the database file that has been selected, retrieved and organized to answer the user's problem or request.

Information retrieval refers to the processes of selection, retrieval and organization of the required subset of data from the database file that provides information specific to the user's problem or request.

A novice database user is one who has no prior experience in the use of a computerized database or in computer programming.

A printed database file is a collection of records presented on paper.

An information request is a question or statement requiring appropriate data to be selected, retrieved, and presented in a form useful to the person making the request.

Information retrieval activities involve the decisions and actions resulting in the selection, retrieval and presentation of data from a database file, and the evaluation of presented information. The decisions include specifying selection criteria for the selection of records, specifying the order in which records will be presented, and specifying the fields within a record to be displayed.

Effectiveness of information retrieval is expressed as the ratio of the number of correct decisions made to the total number of decisions required to retrieve the appropriate information.

Efficiency of information retrieval is a composite measure expressed by the formula:

Efficiency = Effectiveness * Time Standard / Time Taken

A detailed description of the measure of efficiency is included in Chapter III.

Evaluation effectiveness is expressed as a ratio of the number of correct decision made to the total number of decisions required when judging the accuracy of retrieved information for a specified information request.

The MTE group participated in a sequence of learning activities for information retrieval strategy development during which they first manually, then subsequently electronically, retrieved information from a database file.

The EO group participated in a sequence of learning activities for information retrieval strategy development during which they electronically retrieved information from a database file.

CHAPTER II

REVIEW OF THE LITERATURE

The literature review addresses issues that affect the development of instructional strategies for improving the effectiveness and efficiency with which novice database users retrieve information from an electronic database. The first section of this review examines the current status and pedagogical advantages of database use and addresses issues affecting successful application of database skills. Topics covered include (a) trends in the use of database technology, (b) integration of database technology into the curriculum, and (c) research devoted to database skills. The second section of this review examines issues affecting the development of instructional materials, treatments, and research methods appropriate for this research. The topics include (a) the acquisition of skills, (b) problem solving, and (c) implications for the design of instructional materials and treatments. Of particular interest in the problem solving literature is information on the role of mental models, as this provides a theoretical foundation for the variation in treatments in the proposed experimental research activities.

Throughout this review, retrieval of information from an electronic database is viewed as a problem solving activity; however, it should be emphasized that the purpose of acquiring the skills and strategies necessary for electronic information retrieval is not an end in itself, but a means of accessing information that will enable teachers and students to participate in more diverse problem solving activities. These uses of database technology in a wide range of curriculum areas reflect also the increasingly varied economic, political, and social applications of database technology in our society.

Trends in the Use of Database Technology in Society

The ubiquitous role of database technology in society was recognized by Oborne (1985), who described offices as "essentially information processing and transformation systems" (p. 317), and by Gaines (1979), who viewed the emergence of database technology as one of the most significant features of the transition in computer applications. The change in emphasis from "data processing" to "information processing" developed simultaneously with the increasing power of computer hardware and software, a

> transition ... of key importance to those of us who see the possibility of far closer interaction between people and computers than is being achieved at present - some form of man-machine symbiosis in which the strengths of each partner mutually compensate for the weaknesses and complement the strengths of the other. (Gaines, 1979, p. 481)

For such a symbiosis to occur, users of information retrieval systems need to understand the capabilities of these systems, and possess the skills to use them to their advantage. Lindamood (1984) predicted that the user will "not only find information technology being used in more activities, he will also find that the systems themselves are more complex in concept and in use, because they do more things" (p. 30-31).

Jagodzinski (1983) recognized the paradox of increasing information processing capabilities and decreasing expertise of system users when he stated that

> a consequence of the falling cost of computer hardware is the rapid growth of interactive information and decision-support systems with an increasing proportion of their users having no experience or training in the use of computers. (p. 215)

The increase in the use of information systems by non-computer professionals was noted also by Paxton and Turner (1984):

> Although computer usage was once the domain of computer professionals, the community of computer users is becoming increasingly heterogeneous, including scientists, managers, educators, students, and administrators... whilst these individuals are not computer professionals, they are interested in utilizing the computer as a tool to facilitate their work and activities. (p. 138)

The trend towards use of information retrieval systems by non-computer professionals has led to increased interest in user-related aspects of computer-human interaction (CHI). In a report on research into the CHI needs of beginning computer users, Paxton and Turner (1984) found that several researchers (Edmonds, 1981; Shneiderman, 1980; James, 1981;

Thomas, 1981; Eason & Damodaran, 1981; Sneeringer, 1978) suggested that user access to database systems could be improved by providing interfaces that are simple to learn and easy to use.

Spence and Apperley (1982) pointed out that the computer-human interaction issues associated with how users retrieve information from large business databases have received very little attention. This is in sharp contrast to the extensive literature available concerning the computer science focus on the design of databases (e.g. Bates, 1986; Croft & Thompson, 1987; Soergel, 1982), and the vast research conducted in library science in information retrieval from online catalogs, bibliographic databases, or document retrieval systems (e.g., Derr, 1984; Harter & Peters, 1985; Hildreth, 1987; Kibirige, 1988; Oldroyd & Schroder, 1982; Schroder, 1983; Shaw, 1986). While the library science research is useful in extending the skills of library professionals who search large commercial databases and in providing greater insight into the difficulties faced by most library patrons who use electronic information retrieval tools, it provides little guidance for training teachers to incorporate the application of database skills into the curriculum.

Radecki (1983), citing Salton (1979), identified three categories of information retrieval systems. In increasing order of complexity they are (a) data or discrete retrieval systems, (b) document or reference retrieval systems, and

(c) question answering or fact retrieving systems.

The information system being used in this research most closely resembles a data or discrete retrieval system in Radecki's framework. He stated:

> ... in a discrete retrieval system a small number of well-defined attributes (of interest to the system users) is normally used to characterize each object in the collection. Moreover, the system data base can be accessed only through an element of a restricted set of pre-specified types of query representations; these query representations are usually Boolean combinations made up of certain attributes and the logical operators of negation, conjunction and disjunction ... From the point of view of retrieval effectiveness, the responses retrieved by discrete retrieval systems are characterized by the highest possible values of recall and precision ... Lower values of recall and/or precision can only arise as a result of factors such as: inadequately formulated queries, errors in creating the representations of the objects and/or queries, technical and/or methodological failures in the retrieval process (1983, p. 408).

For this research activity, reduced effectiveness in information retrieval (lower values of recall or precision) can be attributed to ineffective use of skills which can be isolated and identified.

The Integration of Database Technology into the Curriculum

Bolt (1986), Collis (1988), Troutman and White (1988) and Rawitsch (1988), recognized that the educational advantages of including database activities in the curriculum cannot be achieved until teachers use this technology.

Bracey (1984) claimed that "there are fundamental educational concepts implicit in allowing children the opportunity to use a data retrieval system. Inherent are ideas of classification, ordering, sorting and the forming of logical conclusions from data" (p. 100). Similarly, Rooze (1986) believes that database activities enable students to find relationships, identify cause and effect, compare and contrast data drawn from different sources, draw conclusions based on data, and evaluate those conclusions.

Johnston (1985) viewed the ability to obtain information from databases as an important academic skill, while Parisi (1985) believed that database skills are fast becoming essential for effective citizenship, and that this is a compelling rationale for using database activities in the social studies curriculum. Students who retrieve information from and manage databases develop skills in their ability to identify information needs, generate problem statements, retrieve relevant information and design strategies for organizing data (Parisi, 1985).

The role of database experiences in helping students develop skills in thinking and problem solving, particularly within the social studies curriculum, has been highlighted in the literature (e.g. Lengel, 1987; Pon, 1984; Rooze, 1983; Sartor, 1986), and suggestions for incorporating database activities in the curriculum have abounded (e.g. Brown, 1987; Hannah, 1987; Hunter, 1985; Little, 1986; Martin, 1984; Pruitt & Dowling, 1987; Resnik, 1984; Thompson & Vaughan, 1986; Troutman & White, 1988; Weaver, 1986). The availability of many commercial database packages for instructional use provides teachers with opportunities to incorporate database activities in their instruction. However, Rooze (1987) emphasized that the activities of the teacher in the planning and guiding of database activities, and the creation of an environment that encourages effective questioning and thinking, determines the value that can be gained from the database activity. As is the case with other micro-computer-based educational activities, the availability of instructional database activities is relatively recent, and the majority of teachers have little or no training in their use.

Research into Database Skills

Research studies of the information retrieval skills of novice users of discrete data information systems have been limited in number and scope, and none of these studies has addressed issues related to methods of training that could be used to assist teachers gain database skills.

Brosey and Shneiderman (1978) examined the effect of relational and hierarchical data models on the comprehension, problem solving, and memorization performance of undergraduate students. Although the subjects were novice database users, the beginners' group had completed between two and three terms of computer programming, and the advanced group had completed six or more terms of computer programming. Significant differences between group performances were observed for the data model used, presentation order, subject background and tasks. Hierarchical data structures resulted in higher performance levels than relational data structures, although this difference was less marked for advanced users than beginners. Beginners showed the greatest improvement in performance due to the presentation order effect, suggesting that once the beginners had some practice and understood the task, they were able to perform at approximately the same level as the advanced group. An important limitation of this study was that data structures were represented on paper and all data collection occurred off-line, thus it is hard to see the implications for training teachers who have virtually no computer proficiency.

Recognizing that an increasing number of "casual users" access databases, Ray (1985) examined the effect of different data models on casual users' performance and confidence in writing database queries. Three experimental groups were each presented with one data model representing either a relational, hierarchical, or network data structure. Subjects were required to write English-like queries that would result in the retrieval of specific information and to assess their level of confidence that the written query would provide the information needed. Written queries were graded for (a) the specification portion, (b) the condition portion, and (c) the navigation portion. Subjects using the relational model performed significantly better than the group using the hierarchical model when writing the specification portion ($p<0.05$) and the navigation portion ($p<0.01$) of the queries. This is in contrast to Brosey and Schneiderman (1978) who had found that memorization and comprehension of the data structure were significantly better for subjects using the hierarchical model rather than the relational model.

Elkerton (1985) conducted a study in which three levels of online command selection aids, each having three types of dialogue--user-initiated, computer-initiated and mixed-initiated--were developed and evaluated to help novice users in command selection during information retrieval activities. Information retrieval performance, strategy, and subjective evaluations were collected from the aided users to determine which model and dialogue were least intrusive and possibly most beneficial when provided to novices. Performance and strategy were then monitored in a final unaided session to determine the amounts of transfer of training in each of the three models.

Elkerton found that online command selection aiding improved the performance of slower novices without interfering with the faster novices. The computer-initiated dialogue was least favored by subjects in each of the three groups, and there was a significant difference between subjects' preference for either user-initiated or mixed-initiated over the computer-initiated dialogue. However, other results reported in this study are difficult to interpret, because the researcher chose to compare the performance results for the three heterogeneous treatment groups with the performance of less heterogeneous control groups--slow novices and fast novices.

White (1985) conducted a study using intact classes of high school students to address the question: "Can the manipulation of a computerized file management program contribute to positive gains in students' ability to locate, gather, organize, and evaluate information?" (p. 1). Although recognizing the importance of structured experiences in skill development (p. 9-10), the treatment groups in White's study not only differed in the medium in which the data were presented--computerized versus card--but also in the degree of structure provided for the manipulation of the data. Although results indicated a significant difference in performance between the two treatment groups ($p<0.05$), it is quite possible that this was the result of the interaction of structured activities with the treatment medium, rather than only the result of using a computerized file management program.

Katzeff (1986) generated Category And Restriction (CAR) as a query language for online use by novice computer users in posing retrieval questions to a computer system. CAR limited the subjects to defining the category--persons, activities, or places--and the restrictions--no restrictions, one restriction with or without NOT, two restrictions with or without AND, OR or NOT. Twenty questions were posed using ten logical forms; each logical form was repeated twice within the question set. The most striking finding was that subjects seemed to prefer the use of one complex query to two simple queries when dealing with complex questions, although it required more problem solving activity. Katzeff inferred from the think-aloud data that variations in performances were a consequence of the differing mental models used by the subjects. The spontaneous generation of additional search queries by some subjects was apparently the result of their developing mental models with greater predictive power.

Katzeff's study was limited by the query language to query situations that were relatively simple; retrieval of multiple categories using more than two restrictions is more representative of real world applications. However, even in her limited environment, the students' performance ranged from 96 percent for the less complex questions to 45 percent for the more complex questions, indicating that the subjects experienced difficulty with information retrieval problems as their complexity increased.

In a later study, Katzeff (1988) investigated the effect of different conceptual models upon reasoning in a database query writing task. Four experimental groups were used; subjects in these groups were provided with a manual in conjunction with four conceptual model conditions. Performance on the query writing task was higher for the two groups who received either of the two "set" models.

Although one of the set models provided a deeper logical analysis than the other, the subjects receiving this model did not produce a larger number of correct queries.

The importance of viewing database query writing as a problem solving activity was emphasized by Katzeff (1988). She claimed that as a problem solving activity the process should not be viewed as completed until a computer response has been obtained and that response evaluated by the user.

The Acquisition of Skills

As previous research has not addressed the acquisition of database concepts and skills by teachers, approaches to skill acquisition need to be examined.

From Novice to Expert

The concept of expertise is of interest to researchers because it represents the ultimate level of skill proficiency. However, it is unrealistic to expect that expertise can be gained easily or quickly. Consequently, the study of the transition from the status of novice to expert provides researchers with valuable information on how skills are acquired, and suggestions for providing instruction in skill acquisition.

Expertise has been defined by Johnson, Zualkernan and Garber (1987) as

> a kind of operative knowledge. It is characterized by

> generativity, or the ability to act in new situations, and by power, or the capacity to achieve problem solutions. Expertise is a kind of knowledge, and not a property of the behavior we observe as individuals perform tasks" (p. 162).

Kolodner (1983) viewed expertise similarly when she observed that there are two major differences that distinguish novices from experts. The first difference is that experts are more knowledgeable about their domain, and the second is that experts know how to apply their knowledge more effectively than novices. In the area of Artificial Intelligence, the emphasis has been on the representation of knowledge (Feigenbaum, 1977, cited in Kolodner, 1983; Frenzel, 1987) and the extraction of the rules that experts use in decision making (Frenzel, 1987; Shortliffe, 1976, cited in Kolodner, 1983; Waterman, 1986), but little attention has been paid to the way in which individuals change the use of their experience as they move from novice to expert. Kolodner claimed that

> the evolution from novice to expert requires introspection and examination of the knowledge used in solving problems. Thus, we see human experts interpreting new cases in terms of something with which they are already familiar (either a previous case or generalized knowledge). This implies that as an expert is having new experiences, he is evaluating and understanding them in terms of previous ones. In the process, he must also be integrating the new experience into his memory so that it too will be accessible to use in understanding a later case (p. 501).

The transition from novice to expert was examined by Dreyfus and Dreyfus (1984, 1986), who studied the skill-acquisition processes of airplane pilots, automobile drivers, and adult learners of a second language. They observed a common pattern in all cases, which they described as a five stage model of skill acquisition:

1. Novice. The beginner learns context-free rules or facts (such as the relative value of chess pieces, and the types of moves each piece can make).

2. Advanced Beginner. The learner gains experience in real situations, recognizing features which make application of facts or rules appropriate.

3. Competence. As experience increases, the context becomes more complex, and the learner adopts a hierarchical view of decision-making. This involves detached planning, conscious assessment of the environment, analytical rule-guided choice of action, followed by critical evaluation of the outcome.

4. Proficiency. Further experience may lead to intuitive "seeing" of an appropriate plan, thus making unnecessary the detached planning and conscious assessment of the environment required in the stage below. However, at this level, learners may still need to focus on the actions constituting the intuitive plan (i.e. they see what needs to be done, but decide how to do it).

5. Expertise. At this level, the expert performer understands, acts, and learns from results without any conscious awareness of the process.

Dreyfus and Dreyfus (1986), in describing this pattern of skill acquisition, claimed that

> as human beings acquire a skill through instruction and experience, they do not appear to leap suddenly from rule-guided "knowing that" to experienced-based

> know-how. A careful study of the skill-acquisition process shows that a person usually passes through at least five qualitatively different perceptions of his task and/or mode of decision-making as his skill improves. (p. 19)

Although this five stage model of skill acquisition describes the progress from novice to expert, Dreyfus and Dreyfus (1986) do not claim that all people will achieve an expert level in all skills, or that performance at any particular level will be uniform within a group. They consider that the importance of the model is twofold:

1. When individuals are confronted with a particular type of situation in their skill domain, they will usually approach it first in the manner of the novice, then of the advanced beginner, and so on through the five stages.

2. The most talented individuals employing the kind of thinking that characterizes a certain stage will perform more skillfully than the most talented individuals at an earlier stage in the model.

It seems reasonable to assume that although the development of expertise is generally achieved through extensive experience over a long period of time, it could be expected that effective instruction and relevant experience could assist the learner in reaching the level of competence--Stage 3 in the skill acquisition model--in a shorter period of time.

Wiedenbeck (1985) reported a three stage model of skill development that had been proposed by Fitts and Posner (1967), and Anderson (1982):

1. The cognitive stage. The learner studies or is instructed in the skill to be acquired. This results in a knowledge of the rules and sequences of activities that go into the skill, but not a smooth procedure for performing it.

2. The associative stage. An efficient procedure for performance becomes established, speed increases, and errors in performance of the skill decrease. At this stage, the declarative knowledge becomes less important to everyday performance, and is only needed when unusual conditions or errors arise.

3. The autonomous stage. Performance is characterized by high speed and accuracy, and the existence of a set procedure for performance of the skill. Declarative knowledge is rarely used at this stage, and may be difficult to access. Automatization of the skill may make it possible to perform the skill while simultaneously performing some other attention-demanding task.

Using the 3-stage model as the basis for two studies of the differences between novices and experts in programming subtasks, Wiedenbeck found that experts were significantly faster and more accurate than novices in recognizing syntax errors in program segments and understanding the structure and function of coded program segments.

Both the Dreyfus and Wiedenbeck models of skill acquisition include an initial period of learning through instruction later followed by more effective, efficient performance in a complex problem solving situation. Using these models, measures of effectiveness and efficiency can be used as measures of relative proficiency.

The Role of Experience

Dreyfus and Dreyfus (1986) emphasized the importance of concrete experience when they claimed, "someone at a particular stage of skill acquisition can always imitate the thought processes characteristic of a higher stage but will perform badly when lacking practice and concrete experience" (p.35). Best performance in a skilled task results from the intuitive use of similarity and experience, ultimately leading to performance at the expert level.

The role of experience in the appropriate mapping of concepts applicable in the problem domain was investigated by Cooke and Schvanelveld (1988) who analyzed the cognitive organization of a set of abstract programming concepts. Four groups of programmers (naive, novice, intermediate and advanced) were required to rate the relatedness of 120 pairs of concepts on a scale from zero (highly unrelated) to nine (highly related). Distinctively different semantic networks were generated for each of the four groups, representing four phases from naive knowledge to expert knowledge of those programming concepts. These varying concept maps, together with the results of analysis of the concepts misdefined at each level, demonstrated that increasing experience leads to more appropriate concept mapping, and provided the researchers with guidelines for designing programmer training and for the representation of expert tacit knowledge.

The transition from novice to experienced computer user was viewed by Allwood (1986) as a process in which difficulties faced by novices were overcome.

1. Novice users had to develop and debug routines to use the computer whereas experienced users executed the same task as a routine skill.

2. Novice users had varied and error-filled patterns of knowledge about computers, resulting in poor access to declarative knowledge in the task situation, ineffective classification of task problems, and ineffectiveness in detecting the source of their errors.

3. Novice computer users were negatively affected by a change in context between the learning and recall task.

4. Novice computer users usually lacked an integrated model of the computer system and had difficulty in drawing correct conclusions about the properties of the system and the cause of errors.

5. Novice computer users lacked an understanding of the ways in which the steps in a task combined to achieve a required goal.

Allwood (1986) emphasized the need to investigate "how novices use general and specific problem-solving heuristics ... [and] the importance and function of both conceptual understanding and of routinized knowledge in novices' computer interaction" (p. 653).

The role of prior experience in non-computer activities in assisting the novice computer user in acquiring computer based skills was addressed by Oborne (1985), who suggested that "the design of data bases should relate in some way to the normal organization of information within an office". Oborne's suggestion that a familiar model would facilitate later information retrieval implies that application of skills in unfamiliar environments can be facilitated by prior experience in analogous task environments.

The method by which effective application of skills occurs is another aspect requiring investigation. Landa (1983) called operations that transform material objects motor operations, but stated that "transformations we perform on real material objects can be performed on their mental images as well" (Landa, 1983, p. 168). These transformations of images he called cognitive operations. Motor operations (or cognitive operations which have motor prototypes) can be taught by both prescription (a statement of methods) and demonstration; the importance of demonstration or concrete experience of the operation as a learning experience was emphasized by Landa when he stated, "if you neither know the individual characteristics of students nor can effect individualized instruction, then use a combination of prescriptions and demonstrations when possible" (p. 194). Gray and Orasanu (1987, p. 203) reported that in research conducted by Kieras and Bovair (1984) "subjects who were taught [a] mental model were able to learn procedures faster, retain them better, and execute them faster than subjects not taught the mental model". Brooks and Dansereau (1987) reported that Mayer (1976) found that beginning computer programmers in the model condition excelled on learning problems requiring interpretation, while nonmodel subjects did better on tasks requiring program generation similar to the tasks given during instruction. The effect of concrete experience in increasing learner understanding was observed by Shute (1979), while Marzola (1987) emphasized the need to form "connections" between actions on objects and the concepts they illustrate through the use of verbal mediation and symbolic notation.

The effect of subjects' concrete experiences in manually retrieving information from a printed database will be examined in this research. If concrete experiences facilitate the development or clarification of concepts and/or mental models, and if these concepts and models promote applications of skills, it should be possible to observe increases in effectiveness and efficiency as subjects retrieve information from an electronic database and critically evaluate the retrieved information.

Problem Solving

If retrieving information from a database is viewed as a problem solving activity, the design of appropriate instruction should draw on the findings from the problem solving literature. Areas of particular interest to this research include (a) mental representations during problem solving, (b) approaches to problem solving, (c) algorithms and heuristics, and (d) understanding, interpretation and representation of the problem.

Mental Representations During Problem Solving

Johnson-Laird (1985) described three types of mental representations. The first of these three types, propositional representation, are strings of symbols corresponding to natural language. The second type, mental models, are structural analogies of the world, while the third type, images, are perceptual correlates of models from a particular view. Johnson-Laird suggested that mental models make it possible for the subject to reason without using formal logic. The literature on the role of imagery, metaphors and analogies in the development of conceptual and mental models provides insight into the importance of providing experiences and models during instruction that promote the use of natural reasoning practices.

Imagery. Kaufmann (1980) reported that in "complex problem-solving activities ... there is reason to believe that productive thinking and imagery are intimately related" (p. 12). The role of imagery in problem solving was thought by Kaufmann to be the medium through which the problem solver moves towards a problem statement or a goal statement. The importance of the concreteness of stimuli and the resulting use of imagery in abstract tasks was described by Kaufmann (1980) when he reported that

> imagery is conceived by Paivio to be linked to the concrete aspects of a situation. The more concrete the stimulus in the task situation, the more easily imaginal processes are elicited which are functional as mediators in these kinds of situations. The verbal processes are regarded as being less connected to the concreteness of the task-situation, and more functional as mediators in abstract tasks. Imagery is, further, regarded as a dynamic system, promoting flexibility and speed in the mediating process, whereas the verbal system is regarded as having a more static, labelling function. (p. 37)

When discussing the role of visual imagery in transformational activities necessary for dealing with unfamiliar tasks, Kaufmann (1980) supported the view held by Bruner (1966) that language is "the most appropriate medium for effectuating transformations (as well as other activities) as such" (p. 118). Kaufmann claimed that through effective use of language "operations may be handled much more economically and rapidly than is possible through the more awkward and laborious medium of visual imagery" (p. 118). However, Kaufmann believed that Bruner had not fully considered the implications of the "enormous information condensation" that has to take place if effective use of language is to occur. Kaufmann proposed that in contrast to language, imagery is more idiosyncratic, varied and flexible. For these reasons, Kaufmann viewed imagery as

> a representational system especially adaptable for the transformational activity needed in tackling tasks which possess a high degree of novelty, and which resist being handled solely through the application of general principles and rules. (p. 119)

In the context of information retrieval from an electronic database, it is most likely that novice information retrievers would view each request for information as a task possessing a high degree of novelty. Further experience in information retrieval would reduce the degree of novelty, and it is likely that a shift from visual to verbal to automatic processing would occur as the information retriever gained expertise through greater experience.

The relationship between visual imagery and concrete stimuli was also discussed by Greeno, James, DaPolito and Polson (1978), who reported that Humphreys and Yuille (1971) found that in paired-associate learning situations with concrete stimuli, learning occurred in an all-or-none fashion. One might infer that concrete stimuli, once stored as representations, are easy to retrieve and further strengthening of the association is not needed. On this interpretation, items with concrete stimuli had only one stage of learning because the second stage occurred with probability close to 1.0 when the first stage was accomplished. If stimulus concreteness has a strong effect on the efficiency with which subjects can generate appropriate associations, it seems reasonable to assume that maximizing stimulus concreteness in electronic information retrieval activities through prior concrete experiences could ease the retrieval of relevant ideas.

Conceptual models and mental models. Young (1981) reported that ideas concerning mental models were discussed as early as 1943 when Craik (1943) suggested that "we build internal models to represent, and to some extent mimic, objects in the physical environment, thereby enabling us to predict and deal with events in the outside world" (Young 1981, p. 51).

Carroll and Mack (1985) emphasized the differences between models and metaphors by stating

> the distinction ... resides chiefly in the open-endedness, incompleteness, and inconsistent validity of metaphoric comparisons versus the explicitness, comprehensiveness and validity of models. ... A model is a description of a target domain or object that seeks to faithfully represent the actual elements, relations and mechanisms that are constitutive of objects in that domain. ... While models are designed to represent some target domain, metaphors are chosen or designed to invite comparisons and implications which are not literally true. (p. 53)

They concluded that an effective model is generated by the learner when appropriate metaphors, accompanied by extensive hypothesis testing by the learner, occurs.

Borgman (1986) reported that

> a small but growing group of researchers is experimenting with a theory that attempts to explain part of the cognitive processing involved in reasoning about computers. The theory is that of "mental models". At the present stage of research in human-

> computer interaction, "mental model" is a general concept used to describe a cognitive mechanism for representing and making inferences about a system or problem which users builds as they interact with and learn about the system. The mental model represents the structure and internal relationships of the system and aids the user in understanding it, making inferences about it, and predicting the system's behavior in future instances. The model is thought to be dynamic, in that it can be "run" to test hypotheses about the mechanism's behavior. When the model is appropriate, it can be helpful (or perhaps even necessary) for dealing with the device" (p. 47-48).

However Borgman (1986) and Young (1981) also emphasized that inappropriate or inadequate models can be developed and easily lead to misconceptions and errors.

Jagodzinski (1983) claimed that novice computer users require a conceptual model of the computer system so that they can form a clear idea of what the system is doing and what it can do, and that computer system designers rely heavily on the user's existing knowledge of a pre-computer system to help their formulation of a representation of the computer system. "In this way simple analogies can be drawn, for example between Kardex files and computer files, invoices and VDU screen 'forms'" (p. 219). However, difficulties arise when the computer system performs completely new functions that were not present in the old system or when users have had "no pre-computer experience of this particular part of the real world" (p. 219). As the use of computerized information processing systems increased, the opportunity for experiencing equivalent non-computerized systems decreased. Consequently, many young adults lack the relevant pre-computer experience assumed by systems designers and instruction in the use of these systems may need to include experiences that the user can then incorporate into a mental model.

Terminology associated with descriptions of user models in the literature is at times ambiguous. Young (1983) identified eight types of models that could be described as being "conceptual models" or mental models. Norman (1983) distinguished between "mental models" and "conceptual models" by classifying models which users have in their minds and that are not observable as "mental models"; in contrast to this, "conceptual models" were defined as models provided to the user by a designer, trainer or experimenter as the basis for users to build their own mental model. In this study, concrete experience in manually retrieving information from a printed database is an implicit conceptual model from which subjects may then formulate a mental model.

Moran (1981) and Norman (1981) questioned whether people would make the effort to develop their own models of a complex system thus leading Borgman (1986) to propose the hypothesis that

> if it is true that people can use the system better if they have a correct mental model of it, then those trained with a conceptual model should perform better than those trained with procedural, or step-by-step, instructions that do not provide a conceptual framework for system operation. (p.48)

Indirect testing of Borgman's hypothesis in computer applications has occurred in the research of du Boulay and O'Shea (1978), du Boulay, O'Shea and Monk (1981), and

Coombs, Gibson and Alty (1982) on computer programming, Young (1981), Bayman and Mayer (1983), Halasz and Moran (1983) and Halasz (1984) on calculator use, and Foss, Rosson and Smith (1982), and Douglas (1983) on text editing. Du Boulay and O'Shea, du Boulay, O'Shea and Monk, and Coombs, Gibson and Alty found that computer programming performance is more effective when novice programmers have a functional view of the computer from the perspective of the programmer. Bayman and Mayer, and Foss, Rosson and Smith found that training involving a conceptual model was superior to procedural training alone. Halasz found that the positive effect of training using a conceptual model was only applicable to complex tasks in calculator use requiring problem solving behavior. The implications of these findings for this proposed study are that instruction incorporating both procedural training and a conceptual model would likely be more effective than training incorporating just procedural training alone.

In contrast to these findings, Douglas found that training for text editing where subjects already had knowledge of typewriters actually interfered with their ability to learn text editing. This provides an example of a mental model which is ineffective, resulting from the mapping of inappropriate characteristics from the typewriter to the text editor. To optimize the potential benefit of instruction, the conceptual model should provide the same system products and should also model the same processes as the abstract system being modelled.

In studies of strategy use by people in a variety of tasks, Moray (1978) found that the more effectively the person's mental model functions as a predictive model the more effectively the operator can select appropriate strategies.

Researchers (Gentner, 1981; Di Sessa, 1981; Eisenstadt, Laubsch and Kahney, 1981; Norman, 1983; Kahney and Eisenstadt, 1982) emphasized the importance of the interaction of general world knowledge with domain specific knowledge in the construction and evaluation of effective predictive mental models. This interaction was observed by Kahney (1983), who noted that there is an important distinction between "talented" and "average" novice programmers. They are distinguished on the basis of the number of programming concepts with which they are familiar, and the degree to which the familiar concepts have been understood and applied. Kahney noted that talented novices have "much more in common with experts in terms of the way their knowledge is organized and in the way in which it is brought to bear in solving programming problems" (p. 127). The talented novice develops and uses mental models to understand the way in which a particular effect is achieved, whereas the average novice commits a segment of code to memory together with the rule for when the code should be used without gaining an understanding of the way in which the required effect is achieved.

Approaches to Problem Solving

Kahney (1983) examined the mental processes of novice computer users when transforming a verbal statement of a problem into a programmed solution to the problem. He concluded that "problem solving in such a task involves understanding what the problem is, finding or devising an algorithm, and writing and debugging code" (p. 122). The three phases are, Kahney claims, characteristic of the "modal model" of problem solving, described more generically as the understanding phase, the method finding phase and the solution phase.

During the understanding phase, two subprocesses are thought to occur. Hayes and Simon (1974, cited in Kahney, 1983) referred to these two subprocesses as "Language Understanding" and "Model Construction". Kahney proposed these processes produce a "problem space containing the initial and goal states of a problem, the problem objects and their properties and relations, and, finally, the operators for transforming the initial into the goal state, plus any restrictions on the use of the operators" (p.124).

Chi, Feltovitch and Glaser (1980) investigated the cognitive structures which novices and experts had acquired in physics and the way in which this knowledge was utilized when both groups were asked to solve physics problems. They showed that although both experts and novices categorize the problem as one of the first steps in problem solving, the eventual effectiveness and efficiency with which a problem is solved depends upon the categorization or schema selection from among competing candidate schemas. Experts used schemas which were organized in terms of physics principles such as conservation of momentum, conservation of energy, or force laws whereas the novices used schemas which were categorized on the basis of objects or surface features of the problems such as ropes and inclined planes with pointers to a range of equation formulae; the relevance of these formulae had to be evaluated before problem solution could occur.

During the method finding phase, domain knowledge is accessed which acts as a link between understanding and solution processes. Access to rules, algorithms and heuristics during this method finding phase is thought to provide direct information about the solution method which can then be applied in the third phase of problem solving.

Problem solving involves the application of the solution methods derived during the second phase. If the problem is not solved in this solution phase, the problem understanding mechanisms are invoked once again, commencing a second iteration through the problem solving phases. Successive iterations may be required before a successful problem solution is obtained.

The stages in problem solving were also addressed by Green, Payne and van der Veer (1983), who indicated that the question commonly asked by researchers interested in the interaction between computers and computer users is "What sort of thinking do computer systems demand?" (p. 7). They suggested that computer-based activities consist of three phases. These three phases are (a) problem definition, (b) algorithm design, and (c) coding of the necessary commands. These phases correspond to the first three steps in Polya's (1945) problem solving strategy, which were (a) understand the problem, (b) devise a plan of action, (c) carry out the plan, and (d) examine the solution.

Algorithms and Heuristics

Problem solving requires the subject to rely on experience as the foundation for developing appropriate problem solving strategies. Lindsay and Norman (1972) described three types of relevant experience that are useful in problem solving. They are:

1. Facts which are immediately available to the subject,

2. Algorithms which are sets of rules that automatically generate the correct answers, and

3. Heuristics which are rules of thumb or general plans of actions.

In discussing the difference between algorithms and heuristics, Mayer (1977) explained that

> algorithms guarantee correct problem solutions since all that is needed is to apply a past set of rules to a new situation. The set of rules can be stored as a subroutine, thus saving memory load. Heuristics on the

PLEASE NOTE:

Poor print on pages 46-55
Not available from author
or school. Film as Received

UMI

other hand, may not solve the problem. (p. 139)

Algorithms were defined by Shute (1979) as procedures for solving structured problems. "They utilize a fixed number of steps and will always result in the same answer when applied to a problem. The use of algorithms does not require abstract reasoning ability" (p. x).

Resnick (1987) described three heuristics commonly used during problem solving. These heuristics are subgoal formation, means-end analysis, and generate-and-test routines. Subgoal formation is used by the problem solver to break the problem down into segments, where the achievement of the subgoal moves the problem solver closer to the final goal. Means-end analysis requires problem solvers to compare their final goals with the results that would be produced if one or more procedures are used. Generate-and-test routines are used by problem solvers when they wish to test the results of specific actions or procedures against the constraints of the problem context.

Each of the heuristics referred to by Resnick (1987) is applicable to information retrieval from a database. Subgoal formation allows the information retriever to focus attention on individual procedures in the information retrieval task. Using this researcher's model, the subgoals could be the definition of sort criteria, the generation of formal selection criteria, the specification of the essential display fields, and the specification of validating display fields. Subgoal formation requires the problem solver to consciously plan, implement, and evaluate procedures related to the subgoal and to remain aware of the general problem task.

The use of a means-ends analysis allows information retrievers to identify the procedures that they think need to be used, predict the results of those procedures and compare those results with the intended goal. The means-ends analysis requires the problem solver to use a mental model of the information retrieval procedures so that accurate prediction of the procedures outcome can be made, and as a means of matching the predicted outcome with the problem goal.

Generate-and-test routines can be used by those information retrievers who have not developed a dynamic mental model or by those who need to extend or verify their existing models during the development of heuristics and specific algorithms.

Retrieval of information from a database file is a problem solving activity requiring both algorithmic and heuristic strategies. The acquisition and application of skills, such as those required for information retrieval, is addressed by the algo-heuristic theory of learning (Landa, 1983) and performance, a theory that

> deals primarily with understanding and describing specific processes--operations and their systems--that turn knowledge into skills and abilities, underlie the latter and the performances that realize them, and allow a person not just to know something but to apply this knowledge to solving problems and performing certain activities (motor and/or cognitive). The algo-heuristic theory of instruction deals with the problem of how to use the information about these operations in

order to purposefully and effectively develop them in the course of instruction. (p. 171)

Landa's (1983) steps involved in building and testing models of cognitive processes include:

1. identifying the elementary operations that are the components of the complex cognitive process;

2. generating a descriptive algorithmic or heuristic model for the complex cognitive process; and

3. verifying the adequacy and completeness of the model by observing each subject's performance when confronted with problems requiring application of the proposed model.

Analysis of information retrieval activities results in the recognition that any descriptive model of information retrieval processes would be part algorithmic, and part heuristic. Consequently, instructional materials and experiences in information retrieval could be based upon a descriptive algo-heuristic model.

Understanding, Interpretation and Representation of the Problem

Interest in the understanding, interpretation and representation of problems has led to research in several areas. Weiser and Shertz (1983) replicated a study conducted by Chi, Feltovich and Glaser (1980) in which the interpretation and categorization of problems by novices and experts were examined. Whereas Chi, Feltovich and Glaser worked in the subject domain of physics, Weiser and Shertz used a set of problems from computer science involving (a) three different application areas, (b) three different data structures, and (c) three programming algorithms. Their findings were similar to those of Chi, Feltovich and Glaser in that they found that expert programmers used fundamental principles such as algorithm type as a category for representing problems, whereas novices favored categorization by the surface features of application area. Experts took more time than novices, presumably because of the greater transformation that was necessary from the surface features of each problem to the experts' internal representation of the problem.

Berger and Wilde's (1984) study of the categorization of algebra word problems by groups of novices (first year algebra students) and experts (analytical geometry students) supported those of Chi, Feltovich and Glaser (1981), and Weiser and Shertz (1983) who found that novices performed at chance levels when categorizing algebra word problems, whereas experts performed very well on this task.

Newell and Simon (1972) regarded the problem solving process as occurring within a problem space; within a problem space, the representation of the problem defines the boundaries of possible solutions to the problem. The importance of the representation of the problem within a problem space is emphasized by Newell and Simon (1972) and Montgomery and Allwood (1978), because they believed that problems which are misinterpreted, and therefor misrepresented, in essential details have little possibility of being solved.

Dunlap and McKnight (1980) identified fourteen steps for solving word problems in mathematics, the first seven of which--perceives, decodes, formulates general meaning, generalizes, translates, determines question(s), and gathers data--are related to understanding or interpreting of the problem. Travis (1981) found that students who could successfully solve algebraic equations experienced equal difficulty in the conversion of word problems to algebraic expressions and equations as other students in the classes.

Interest in the understanding, interpretation and representation of problems led Montgomery and Allwood (1978) to study the representation of statistical word problems by a group of undergraduate university students. They found two fundamental differences between the problem solving approaches of good problem solvers and poor problem solvers:

1. good problem solvers more often "made an exhaustive definition of an essential concept mentioned in the problem" (p. 107) than poor problem solvers, and

2. good problem solvers "appeared to attend more to what was asked for in the problem" (p. 107), whereas poor problem solvers appeared to lack "the ability to eliminate errors made in a particular representation of a problem" (p. 124).

Paige and Simon (1966) investigated the processes that people used in setting up an algebraic equation corresponding to a problem stated in English prose. They found that some subjects were able to directly translate from the prose to an equation, some needed to physically represent the problem using diagrams or other non-formula representations, while some generated intermediate equations which then required further substitution from the problem statement. They claimed that "translating the word problems into equations requires a knowledge of the meanings of words that play a grammatical role, or an arithmetic one, but not the meanings of those that play a substantive role. The substantive terms, or the phrases constructed of them ... are for the most part simply translated into conventional names of variables" (p. 84). These language comprehension skills are then implemented in conjunction with assignment of variable names, selection of symbols for mathematical operations and the positioning of elements within the equation.

This researcher believes that a similarity exists between problem representation in an algebraic formula and the generation of machine-processible statements when retrieving information from an electronic database. Successful information retrieval requires the user to explicitly specify the sort requirements, selection criteria and field display options; the ability to do this in a manner appropriate to the information request requires language comprehension skills, knowledge of the database structure and contents, knowledge of information retrieval processes, and the ability to develop a strategy for selecting appropriate commands to provide the requested information.

Implications for the Design of Instructional Materials and Treatments

The literature reviewed indicates that the following questions need to be considered when developing instructional materials and treatments:

1. What knowledge and skills should be included in the instructional materials?

2. What is the most effective method of instruction?

3. What conditions are required to increase transfer of learning as subjects apply knowledge and skills to novel problems?

4. What measures can be taken to facilitate the transfer of learning when subjects retrieve information from a computerized database file?

Answers to questions one and two are provided in the literature on expertise (Dreyfus & Dreyfus, 1984, 1986; Kolodner, 1983; Wiedenbeck, 1985), and instructional design (Dick & Carey, 1985; Landa, 1983). The facts and rules used by proficient or expert information retrievers need to be identified by completing a task analysis for inclusion in the instructional materials. Methods of instruction, addressed in question two, include the presentation of context-free rules and facts, followed by experiences that enable the learner to gain experience in recognizing when to apply the rules or facts. Where possible, algorithms should be introduced.

The transfer of learning, addressed in questions three and four, is discussed in the literature by many researchers (e.g. Brooks & Dansereau, 1987; Mayer, 1975; Landa, 1983; Gray & Orasanu, 1987; Marzola, 1987; Johnson-Laird, 1985; Kaufmann, 1980; Young, 1981; Borgman, 1986; Jagodzinski, 1983; Norman, 1983). Some researchers focus on the type of transfer, such as (a) content to content transfer, which would be promoted in this research activity by ensuring similarity of database content for both the instructional and the data collection activities, (b) skills to skills transfer, which can be reinforced through the use of algorithmic approaches to the decision-making processes of information retrieval, and (c) content to skills transfer, which is achieved by ensuring that the subjects can successfully complete the set tasks using the knowledge and skills provided during instruction. Transfer of learning is also increased when subjects are encouraged to focus on the similarity between problems, rather than on the surface features, and through the introduction of heuristics--such as subgoal formation, ends-means analysis, and generate and test routines--which can be used during decision-making when algorithmic approaches are not feasible. The role of imagery, conceptual models and mental models in encouraging creative problem solving and promoting transfer of learning appears to be of greater importance when the problem solving has to occur in a new, and abstract, environment. For subjects applying information retrieval skills in a computerized environment, it seems likely that an effective mental model--possibly enhanced by concrete experiences in manually retrieving information from a printed database--would result in improved effectiveness and efficiency of performance.

The instructional materials developed for this investigation were designed to incorporate each of the approaches described above.

The interaction between novice users and the computer when using the computerized database file introduces complexity that could not be provided for in the instructional materials. Du Boulay, O'Shea and Monk (1981) claimed that system simplicity should be designed taking into account three aspects of simplicity. First, functional simplicity is achieved by limiting the set of transactions available to the user. Second, logical simplicity is achieved by ensuring that the problem task can be achieved by the novice using the available transaction set. Third, syntactic simplicity is achieved by ensuring that the rules for written instructions are uniform, easy to remember and easy to implement. The use of an interactive command file as the communication between the novice computer user and the dBase III Plus database file has been incorporated into this research activity as a means of providing system simplicity.

CHAPTER III

METHOD

The use of electronic databases by teachers as tools for information processing and for the development of higher-level thinking skills has been addressed by many educators and curriculum consultants, business leaders and school communities (Rooze, 1987; Collis, 1988; Troutman & White, 1988). This study was designed to address the need for identifying effective instructional methods for teaching electronic information retrieval strategies. The study compared the effectiveness of two instructional techniques on subsequent information retrieval performances of participants who completed a three-stage sequence of learning activities. These instructional techniques introduced one group of participants, the MTE group, to the development of information retrieval strategies using a printed database file before they applied these strategies to the task of electronic information retrieval from a computerized database file; the other group, the EO group, was introduced directly to the task of developing retrieval strategies for electronic information retrieval from a computerized database file.

Eighty-two upper division undergraduate Education majors who were novice computer users participated in this study that used three instructional stages to provide instruction for entry level skills, provide instruction for the development of information retrieval strategies, and provide instruction for evaluating the accuracy of information retrieved from a database. The instructional design for this three-stage sequence of learning activities is illustrated in Figure 1. This figure will aid the reader as features of the instructional design are outlined.

The research questions examined in this study were:

1. Is there a significant difference between measures of manual retrieval effectiveness and measures of electronic retrieval effectiveness during introductory learning activities in retrieval strategy implementation?

2. Is there a significant difference between performance of the MTE group and the EO group on measures of retrieval effectiveness during subsequent information retrieval learning activities in a computerized environment?

3. Is there a significant difference between the MTE group and the EO group on measures of retrieval efficiency during subsequent information retrieval learning activities in a computerized environment?

4. Is there a significant difference between the MTE group and the EO group on measures of information evaluation?

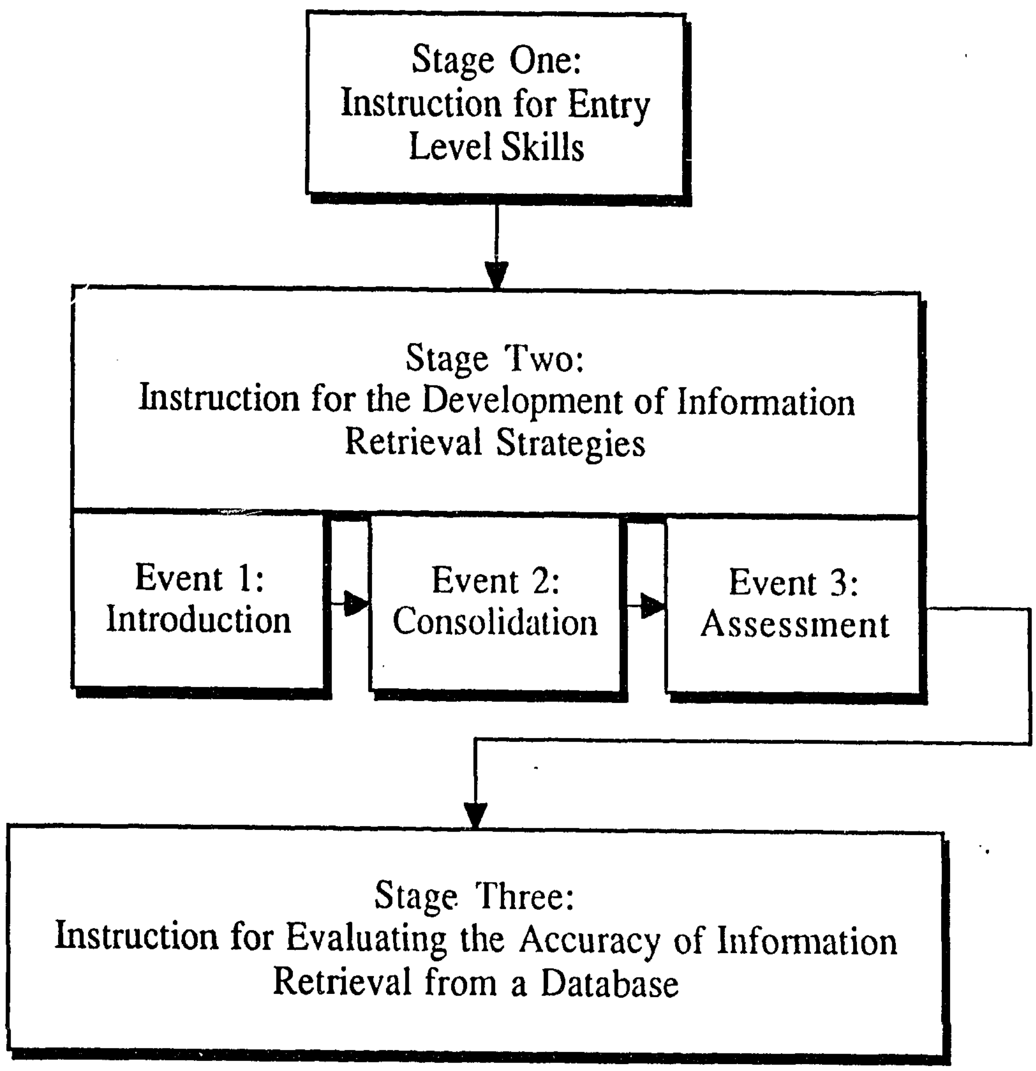

Figure 1. Design of the Three-Stage Instructional Sequence

Instructional Design

Stage One: Instruction for Entry Level Skills

This stage identified a set of entry level skills for the instructional activities in the remaining stages of the study, provided instruction for these entry level skills, and provided criteria for identifying those volunteer study participants who had achieved mastery of these skills.

Materials Development for Stage One Instruction

This stage of instruction began with the identification of the knowledge and skills that are required so that one can gain skill in information retrieval from a computerized database. In terms of the Dreyfus and Dreyfus (1984, 1986) model of skill acquisition, this stage identified a set of context-free rules and facts that help novices progress along the continuum towards expertise. Instructional materials and strategies were then developed to assist database users in gaining those context-free rules and facts as the entry level skills and knowledge needed for their later acquisition of information retrieval skills.

The instructional materials developed for use during Stage One were:

1. a systematically designed instructional package introducing the entry level skills and concepts required for information retrieval from a database, and

2. a criterion-referenced mastery test (CRMT) for the defined instructional objectives in the instructional package.

Instructional package. A systematically designed instructional package, "Database Concepts and Skills", was prepared and field tested by the researcher following the procedures detailed by Dick and Carey (1985). The activities in this instructional package were used as a means of introducing a set of generic concepts and skills required to learn information retrieval strategies from a computerized database. These instructional activities (see Appendix A) provided participants with a series of structured activities that introduced the concepts and skills, tested the participants' knowledge and application of these concepts and skills, and provided performance feedback to the participants.

The following procedures, detailed by Dick and Carey (1985), were completed during the design and development of the "Database Concepts and Skills" instructional package:

1. The instructional goal was identified (see Appendix B).

2. Entry behaviors and characteristics were identified, and an instructional analysis (see Appendix B) of the instructional goal and subordinate skills was completed.

3. Performance objectives for skills identified in the instructional analysis were written (see Appendix B).

4. Criterion-referenced test items, reflecting the behaviors identified in the behavioral objectives, were written. A set of these criterion-referenced test items was used to construct the criterion-referenced mastery test (CRMT) for the instructional package (see Appendix C).

5. An instructional strategy with appropriate instructional materials was developed to produce the learner behaviors specified in the performance objectives.

6. Formative evaluation of all components of the instructional package was conducted. This included validation of the instructional analysis, the performance objectives and the criterion-referenced test items by content experts, as well as one-to-one evaluation, and small-group evaluation with students.

7. Revisions of components in the instructional package were made on the basis of feedback from the formative evaluation process, and from field trialing of the instructional materials.

Criterion-referenced mastery test (CRMT). The CRMT (see Appendix C) contains test items addressing each of the learning objectives for the "Database Concepts and Skills" instructional package. The set of test items in the CRMT was validated by a panel of faculty content experts in information retrieval who matched test items with the behavioral objectives of the instructional package. The validated CRMT operationally defined mastery of the skills and knowledge prerequisite to the following training stage and was used to identify those participants who mastered the learning objectives of the instructional package.

Stage One Instructional Procedures

Participants met with the researcher for two two-hour sessions in the first week of the study during which they completed an informed consent form, the activities in the "Database Concepts and Skills" instructional package, and the CRMT.

The instructional package was included as part of the research activities to help participants acquire a common core of prerequisite information retrieval concepts and skills they would later need to develop effective strategies during either manual or electronic information retrieval activities. Participants spent approximately three and a half hours on the average completing the instructional package.

After completing the instructional materials, participants completed the CRMT. Tests were scored and those who attained mastery of the instructional objectives were identified for data analysis when examining the four research questions during the training stage of this study. Specification of mastery of the instructional objectives as a prerequisite for inclusion of participants' data in this study provided the researcher with a device for strengthening claims to homogeneity of variance within the treatment groups.

Stage Two: Instruction for the Development of Information Retrieval Strategies

Stage Two of the instructional sequence provided three parallel instructional events for each of the two treatment groups. During these activities, participants formulated and implemented retrieval strategies using an algo-heuristic model of information retrieval.

Materials Development for Stage Two Instruction

The instructional materials developed for use during Stage Two were:

1. a set of data, constituting a small database file, that could be used during the instructional activities,

2. three sets of six structured information requests that provided the context for information retrieval activities during instruction,

3. an instructional strategy guide incorporating an algo-heuristic model of information retrieval, and

4. a database user-interface incorporating menu screens to assist novice users, an algo-heuristic model of information retrieval, and performance feedback.

Data in the database. The database consisted of 12 records each containing geographic, demographic, economic and political data concerning one country (see Appendix D). Each record contained 20 fields. The data for the database were taken from The World Almanac and Book of Facts: 1986 (1985). The decision to use 12 records, each containing 20 fields, was made in order to keep the database small enough to facilitate manual retrieval in a printed database environment and large enough to permit the generation of a diverse range of information request.

Information requests. Four parallel sets of information requests were generated. Three sets of these instructional materials were used during the three instructional events during Stage Two, and the fourth set was used during the single instructional event in Stage Three. Each set contained six information requests. Each set was developed as follows:

1. Specifications were developed for each of the six requests. These specifications documented whether or not data needed to be sorted, how many selection criteria were to be defined, the logical operators needed to link selection criteria, and whether or not parentheses were needed to link some selection criteria. A copy of these specifications is included as Appendix E.

2. A set of six information requests was generated, with each request meeting the documented specifications for its respective question. Each request was presented as a simple, conversational statement or question.

Once the first set of requests had been developed, three more sets, parallel to the first were generated by maintaining their structure, format and complexity while substituting alternative selection and/or sort criteria. This was achieved by (a) varying the field(s) involved in the selection criteria, (b) varying the relational operators required for selection criteria, (c) varying the argument(s) within the selection criteria, (d) altering the order in which information needed to be sorted for presentation, or (e) altering the key field by which data needed to be sorted. The four sets and the specifications for each question were judged to be parallel by an independent panel of faculty members experienced in information retrieval.

During Stage Two instruction, participants received the same requests, in the same order, regardless of the environment in which they were working. All requests were presented in writing as part of the strategy guide for each task during the introductory, practice and assessment of retrieval strategies during Stage Two instruction. The four sets of information requests are enclosed as Appendix F.

Instructional strategy guide. The strategy guide (see Appendix G) developed for use during the information retrieval activities was designed for several functions. First, it was designed to guide the participants' decision-making processes by using an algo-heuristic model for devising information retrieval strategies. Second, it was designed to focus participants' attention on one part of the decision-making process at a time. Third, it was designed to help participants implement the decisions recorded on the strategy guide rather than trying to remember what decisions had been made. Fourth, for the Event 1 introductory information retrieval activities using the printed database file, the strategy guide was designed to collect participants' decisions that measured their information retrieval effectiveness.

Database user-interface. A dBase III Plus command file was developed as a database user-interface to guide participants through a series of menu screens where they were required to enter decisions regarding the specification of the sort and selection criteria and to specify the fields to be displayed. (Sample menu screens are included as Appendix H.) The database user-interface (a) required users to implement an algo-heuristic model for retrieval strategy development, (b) accepted and implemented decisions entered via the keyboard so that the participants could retrieve information from the computerized database, (c) provided feedback to participants by displaying both the results of the participants' information retrieval decisions and the correct answers, and (d) recorded participants' decisions, including the time taken to implement them.

Equipment

Computer-based activities in this study were completed using dBase III Plus software (Educational Version) in conjunction with a dBase III Plus command file user-interface and a dBase III Plus database file on an IBM-PC compatible computer. Each participant was assigned a separate command file disk so that only one set of results was recorded on each disk.

Stage Two Instructional Procedures

To initiate Stage Two of the study, all participants responded to the computer-administered demographic data questionnaire. Data gathered included the participants' sex, year of birth, and grade point average. Completing this questionnaire also provided participants with practice in the entering of data by selection of menu options. Subsequently, both the MTE group and the EO group completed the three sets of instructional activities requiring the formulation and application of information retrieval strategies. The three sets of instructional activities provided both a logical instructional sequence and the means by which information retrieval skills could be measured. Participants developed, refined, consolidated and applied information retrieval strategies to complete each set of activities. The components of the instructional sequence were:

1. In Event 1 activities, the MTE group were introduced to the task of manual information retrieval from a printed database file, while the EO group were introduced to the task of electronic information retrieval from a computerized database file. In each case, participants received performance feedback. These activities provided the researcher with measures of retrieval effectiveness in the two retrieval environments that were used to address

research question one.

2. In Event 2 activities, the MTE group completed their first set of retrieval activities in the computerized environment while the EO group completed an additional set in this environment. This event was structured so that the total number of exercises experienced by any participant, regardless of treatment, would be equalized. The activities in this event provided both groups the opportunity to consolidate their information retrieval skills. No data was collected.

3. In Event 3 activities, the MTE group and the EO group applied information retrieval skills, and received performance feedback. Measures of retrieval effectiveness and retrieval efficiency were collected to address research questions two and three.

Table 1 summarizes the sequence, type of activity and environment for each of these activities for both treatment groups.

Table 1

Summary of the Sequence, Type of Retrieval Activity and Database Medium for Instructional Events in Stage Two.

Event	Retrieval Type	Group	Database Medium	Research Question
1	Manual	MTE	Printed	1
	Electronic	EO	Computerized	1
2	Electronic	Both	Computerized	2,3
3	Electronic	Both	Computerized	4

Completion of Event 1 introductory information retrieval activities. Participants in the MTE group were provided with a complete set of data on printed forms. They were instructed to enter decisions regarding their information retrieval strategies onto the printed strategy guide (see Appendix G), retrieve the required information for each information request, and record that information at the lower left hand corner of the printed strategy guide.

The decisions required for the development of a retrieval strategy included:

1. Did the set of data, or the retrieved information, need to be sorted? If so, by which field, and in which order (ascending or descending)?

2. Did the information request include one or more selection criteria? If so, how should the selection criteria

be formally represented?

3. Which fields needed to be displayed to provide the requested information?

4. Which fields, if displayed, were validating fields (fields which would validate the retrieved information in relation to the selection and/or sort criteria)?

After completing the first information request, MTE participants received printed feedback showing the information appropriate to the information request, allowing them to compare their written responses with the correct answer. This procedure was repeated for the MTE participants in each of the six information requests used during Event 1 activities.

In contrast, participants in the EO group were provided with a sample data form, containing a single set of data. They were instructed to enter decisions regarding their information retrieval strategies onto the printed strategy guide, and implement those decisions electronically via the keyboard using the database user-interface.

After completing the activities for each individual information request, EO participants were provided with onscreen performance feedback in the form of a computer generated answer immediately following the display of the information retrieved by the participant's retrieval decisions.

Completion of Event 2 and 3 information retrieval activities. Participants in both groups retrieved information from the computerized database file using the Event 2 and Event 3 sets of information requests specified on the printed strategy guides. For MTE participants, the Event 2 activities provided the first opportunity for them to implement information retrieval decisions electronically. Following each of the Event 2 and Event 3 activities, participants were provided with onscreen performance feedback showing both the information retrieved by their implemented strategy and the information appropriate to the information request.

Stage Three: Instruction for Evaluating the Accuracy of Information Retrieved From a Database

Stage Three of the instructional sequence presented activities that required participants to use an algo-heuristic model of information retrieval to evaluate the accuracy of retrieved information. The measures of evaluation effectiveness were collected during these activities were used to address research question four.

Materials Used During Stage Three Instruction

The instructional materials used during Stage Three were:

1. a set of six information requests parallel in structure to the three sets used during the Stage Two instructional activities (the development of these information requests was detailed previously);

2. a set of six tables of information retrieved in response to the specified information requests; and

3. an instructional evaluation guide, incorporating an algo-heuristic model of information retrieval.

Stage Three Procedures

MTE and EO participants were required to examine tables of information which had been retrieved in response to the specified information requests for Stage Three. Participants were asked to judge whether or not the presented information was appropriate for each information request given the data in the printed database file. For each information request, participants were required to complete an evaluation guide (see Appendix J), where they indicated their assessment of the accuracy of the individual information retrieval decisions. A copy of the Stage 3 information requests and the tables of information that were presented to the participants, the criteria for developing those tables of information and a copy of the Evaluation Guide are included as Appendix J.

Statistical Design

This experimental study investigated the effects of two information retrieval environments on the subsequent performance measures of retrieval effectiveness, retrieval efficiency and evaluation effectiveness. These measures were examined to answer the following research questions:

1. Is there a significant difference between measures of manual retrieval effectiveness and measures of electronic retrieval effectiveness during the introductory activities in retrieval strategy implementation during Stage Two instruction?

2. Is there a significant difference between performance of the MTE group and the EO group on measures of retrieval effectiveness during the Stage Two assessment activities?

3. Is there a significant difference between the MTE group and the EO group on measures of retrieval efficiency during the Stage Two assessment activities?

4. Is there a significant difference between the MTE group and the EO group on measures of information evaluation during the Stage Three information evaluation activities?

Independent Variable

The independent variable in this study was the instructional treatment used to introduce the development of information retrieval strategies. One instructional treatment, manual then electronic (MTE), provided participants with instruction in the development of manual information retrieval strategies before requiring them to apply and refine these strategies for electronic information retrieval. The other instructional treatment, electronic only (EO), provided participants with instruction in the development of electronic information retrieval strategies before requiring them to continue applying and refining these strategies.

Dependent Variables

The five dependent variables measured during the training stage of this study were (a) product retrieval effectiveness, (b) process retrieval effectiveness, (c) product retrieval efficiency, (d) process retrieval efficiency, and (e) evaluation effectiveness.

Two approaches to measuring the effectiveness of information retrieval were incorporated into the data analyses for research question one. The product effectiveness approach used a single binary score to classify the retrieved information for each of the six information requests in a set as being either correct in all respects or incorrect. The process effectiveness approach used a binary scoring system to score each of the decisions made by participants during the information retrieval process for the set of six information requests. The coding criteria for each set of information requests are included in Appendix I. The Stage Two, Event 1 set required 80 decisions, while the Stage Two, Event 3 set required 78. A ratio score for process retrieval effectiveness was calculated by finding the ratio of correct responses to total possible responses for the set.

Information retrieval efficiency was defined by a measure incorporating both process effectiveness and time to completion of the set of six tasks. Information retrieval effectiveness was calculated as follows:

Efficiency = Effectiveness * Time Standard / Time Taken

where

Effectiveness is the retrieval effectiveness ratio score for either product or process effectiveness;

Time standard (measured in seconds) was the mean time, 1335 seconds, required by a group of experienced database users to complete the Stage 2, Event 3 activities with 100 per cent effectiveness; and

Time Taken (measured in seconds) was the time required by the participant to complete the set of information retrieval activities.

Using this formula, information retrieval efficiency for the Stage 2, Event 3 activities was reported as a ratio. For example, if a participant took 1500 seconds to complete the set of activities with an effectiveness score of 0.75, the efficiency score for that participant would be calculated as follows:

$$\text{Efficiency} = .75 * 1335 / 1500$$
$$= .67$$

Evaluation effectiveness was defined as a measure of the accuracy with which participants recognized whether or not decisions leading to the retrieval of information had been made correctly in response to a given information request. The raw score for evaluation effectiveness was determined by coding each of the 60 evaluation decisions for the Stage 3 activities using a binary scoring system. The evaluation effectiveness score for each participant was converted to a ratio score prior to data analysis.

Control of Extraneous Variables

In order to reduce the influence of extraneous variables, the following measures were taken:

1. Participants were novice computer users, as defined in this study.

2. Only data from participants who passed the CRMT were included for data analysis when examining the four research questions.

3. Participants were randomly assigned to treatment groups.

4. Formal protocols were established to ensure that the experience for participants within and between treatment groups was as similar as possible, other than for the intended treatment differences between groups.

5. Participants had completed at least two computer laboratory sessions as part of their instructional computing course prior to commencing the computer-based research activities. This was planned to reduce the differences in measures of efficiency that could have resulted from varied prior experience in keyboard use.

6. Use of a dBase III Plus command file for data collection facilitated recording of information retrieval decisions by using a series of menu screens. In most cases, participants simply had to select a menu option by pressing a number, then pressing the <Enter> key.

8. Participants used the same computer hardware to reduce the likelihood of measures of efficiency being affected by speed of execution of the dBase III Plus command file and identical copies of the same computer software.

Participants

The target population for this study consisted of undergraduate teacher-education students who are novice computer users. For this study, a novice computer user is defined as one who has not learned a programming language, used database or spreadsheet software, or used any form of data management software in a business environment.

The participants in this study were 82 upper division undergraduate education majors enrolled in a compulsory introductory course in instructional computing at the University of South Florida during Fall Semester, 1989. They volunteered to participate in the study, which was completed during the first four weeks of the semester. They received course credit for participating.

Participants were informed prior to the study that the purpose of the study was to investigate the effects of two different database training experiences on participants' subsequent effectiveness and efficiency in information retrieval and evaluation, and that levels of performance during this activity would in no way affect their course grade. Approval to conduct the study using human subjects was granted by the Institutional Review Board at the University of South Florida. Participants were assured of confidentiality of results, and that all data would be recorded and reported in a manner that would preserve anonymity. Written consent was obtained from all participants.

All volunteers were screened by the course instructors who then randomly assigned each to one of two treatment groups, resulting in 41 participants in each group. It was anticipated that the results from some participants would not qualify for use when examining the four research questions for any of the following reasons:

1. Failure to pass the CRMT,

2. Voluntary withdrawal from the research activities, and

3. Failure to complete the full set of activities in the allotted time.

Data were collected for 79 participants who completed the research activities. However, only data for those 46 participants who achieved mastery of the entry level skills and concepts during Stage One instruction, as measured by 80% accuracy on the CRMT, were included in the analyses of the four research questions in this study.

Data Collection During Stage One Activities

Participants' scores on the CRMT were used to identify those participants who achieved mastery of the entry level skills introduced during Stage One of the instructional sequence.

Data Collection and Analysis During Stage Two Activities

Data collection during the Stage Two activities provided the researcher with demographic data for each participant, and measures of retrieval effectiveness and efficiency to answer research questions one through three.

Demographic Data

Participants responded to a computer-administered demographic data questionnaire which gathered data about the participants' sex, year of birth, and grade point average.

Research Question One

Research question one required measures of retrieval effectiveness to be collected during the Stage Two, Event 1 retrieval activities. Data for the MTE group were collected from each participant's completed strategy guide. Data for the EO group were collected by the database user-interface program during strategy implementation. Analyses of variance were used to compare the mean performance of the MTE group and the EO group on measures of process and product information retrieval effectiveness for the Stage Two, Event 1 activities. These analyses were implemented using the SPSS-X procedure ONEWAY.

Research Question Two

Research question two required measures of retrieval effectiveness to be collected during the Stage Two, Event 3 retrieval activities. Data for both the MTE group and the EO group were collected by the database user-interface program during strategy implementation. An analysis of variance with repeated measures on set was implemented using the SPSS-X procedure MANOVA to investigate the difference in process retrieval effectiveness between groups for the Stage Two, Event 1 and Stage Two, Event 3 activities, as well as the difference within groups across sets. An analysis of variance was used to compare the mean performance of the MTE group and the EO group on the measure of product retrieval effectiveness for the Stage Two, Event 3 activities. This analysis was implemented using the SPSS-X procedure ONEWAY.

Research Question Three

Research question three required measures of retrieval efficiency to be collected during the Stage Two, Event 3 retrieval activities. Data for both the MTE group and the EO group were collected by the database user-interface program during strategy implementation. Analyses of variance were used to compare the mean performance of the MTE group and the EO group on measures of process and product information retrieval efficiency for the Stage Two, Event 3 activities. These analyses were implemented using the SPSS-X procedure ONEWAY.

Data Collection During Stage Three Activities

Data collection during the Stage Three activities provided the researcher with a measure of evaluation effectiveness for each participant to answer research question four. Data for both the MTE group and the EO group were recorded on the evaluation guide during strategy evaluation. A one way analysis of variance was conducted to compare the group means for information evaluation effectiveness for the Stage Three activities. This analysis was implemented using the SPSS-X procedure ONEWAY.

Post Hoc Analyses

Following collection of data and the analysis of the four research questions, several post hoc analyses were conducted to provide further information about the nature of the retrieval tasks for novice database users, and to assess the contribution of mastery of entry level skills to retrieval performance.

The post hoc analyses addressed the following questions:

1. What is the degree of relationship between level of mastery of the entry level skills and subsequent measures of information retrieval and evaluation performance?

2. Is information retrieval effectiveness related to the number of decisions required to implement the strategy?

3. Does performance on component tasks during implementation of an information retrieval strategy vary between mastery level groups?

The first question required first-order partial correlations to be calculated, and one-way analyses of variance were conducted to compare measures of retrieval performance between the mastery and nonmastery volunteers in the study. The performance measures analyzed were process retrieval effectiveness during Stage Two introductory activities, process retrieval effectiveness during Stage Two assessment activities, and process retrieval efficiency during Stage Two assessment activities. These analyses were implemented using the SPSS-X procedure ONEWAY.

The second question, investigating the perceived complexity of individual information requests for novice computer users, required a comparison of the process effectiveness with which all mastery participants retrieved information in response to individual information requests during Stage Two assessment activities. These comparisons were accomplished by interpretation of tabulated retrieval performance data collected during the Stage Two assessment activities.

The third question was investigated by comparing the effectiveness with which mastery and nonmastery participants implemented the component tasks during Stage Two assessment activities. The comparisons were accomplished by interpretation of tabulated retrieval performance data, and an analysis of variance which was conducted to compare measures of retrieval effectiveness for each component task between the mastery and nonmastery volunteers in the study. This analysis was implemented using the SPSS-X procedure ONEWAY.

Statistical Analysis Software and Hardware

All quantitative data analyses were completed using SPSS-X, version 3.0, on an IBM 3081K computer at the University of South Florida.

CHAPTER IV

RESULTS

This chapter presents the findings for each of the research questions:

1. Is there a significant difference between measures of manual retrieval effectiveness and measures of electronic retrieval effectiveness during introductory learning activities in retrieval strategy implementation?

2. Is there a significant difference between performance of the MTE group and the EO group on measures of retrieval effectiveness during subsequent information retrieval learning activities in a computerized environment?

3. Is there a significant difference between the MTE group and the EO group on measures of retrieval efficiency during subsequent information retrieval learning activities in a computerized environment?

4. Is there a significant difference between the MTE group and the EO group on measures of information evaluation?

Findings for the following questions addressed through post hoc analysis are also presented:

5. What is the degree of relationship between level of mastery of the entry level skills and subsequent measures of information retrieval and evaluation performance?

6. Is information retrieval effectiveness related to the number of decisions required to implement the strategy?

7. Is performance on component tasks consistent during implementation of an information retrieval strategy?

Stage One: Performance on the CRMT

During this stage of the study, participants received instruction in the entry level skills and completed the CRMT. Of the 79 participants who completed all activities during the study, 46 passed the CRMT. Only data for the mastery participants were included for analysis of the research questions for this study. Table 2 displays the descriptive statistics for the CRMT scores, by group.

Table 2

Descriptive Statistics of Participant CRMT Scores by Mastery Level and Group

Group	n	Mean	SD	Min	Max
Mastery:	46				
MTE	23	87.33	4.37	80.0	97.8
EO	23	88.49	5.36	80.0	97.8
Nonmastery:	33				
MTE	15	71.71	4.57	64.4	77.8
EO	18	67.78	8.73	42.2	77.8

Stage Two: Effectiveness of Information Retrieval During Event 1 Introductory Activities

In Stage Two, Event 1 research question one was addressed: Is there a significant difference between measures of manual retrieval effectiveness and measures of electronic retrieval effectiveness during introductory learning activities in retrieval strategy implementation?

A process effectiveness criterion was applied to score each in a series of decisions made by the participant in response to the individual questions on the decision worksheet for each of the six problems during the Stage Two, Event 1 activities. A product effectiveness criterion was also used to classify the retrieved information for each of these six information requests as being either correct or incorrect. These different evaluation standards permit both a rigorous and a lenient examination of the information retrieval performance.

The obtained process effectiveness ratio scores ranged from 0.42 to 0.97, with a grand mean of 0.70. The process effectiveness mean ratio score for the MTE group was 0.009 lower than the process effectiveness mean ratio score for the EO Group. These descriptive statistics are reported in Table 3.

Table 3

Summary Statistics for Process Retrieval Effectiveness Ratio Scores During Stage Two Introductory Activities.

	n	Mean	Min	Max	SD
MTE group	23	0.696	0.45	0.95	0.154
EO group	23	0.705	0.42	0.97	0.126
Total sample	46	0.700	0.42	0.97	0.139

To determine if there was a significant difference in performance between groups for process retrieval effectiveness, data were subjected to an analysis of variance. The main effect for group was not statistically significant, $F(1,45) = 0.049$, $p > .05$, indicating that there was no significant difference in process retrieval effectiveness performance between the MTE and the EO groups during the Stage Two, Event 1 introductory activities.

The descriptive statistics associated with the ratio scores for product retrieval effectiveness are presented in Table 4. Using the product effectiveness approach, participants' raw performance scores could range from zero through six, and the ratio scores from zero through one. The observed ratio scores ranged from a minimum value of zero to a maximum value of one with a grand mean of 0.529. The product effectiveness mean ratio score for participants in the MTE group was .536 higher than the product effectiveness mean score for participants in the EO Group.

Table 4

Summary Statistics for Product Retrieval Effectiveness Ratio Scores During Stage Two Introductory Activities.

	n	Mean	Min	Max	SD
MTE group	23	0.797	0.50	1.00	0.201
EO group	23	0.261	0.00	0.83	0.250
Total sample	46	0.529	0.00	1.00	0.352

To determine if there was a significant difference between these measures of performance in the two retrieval environments, data were subjected to an analysis of variance. The main effect for environment was statistically significant, $F(1,44) = 64.217$, $p < .001$, indicating that participants who manually retrieved information from a printed database file were significantly more effective in retrieving the required information product than participants who electronically retrieved information from a computerized database file during this introductory set of retrieval strategy activities. These results, which are summarized in Table 5, indicate that MTE participants implemented accurate retrieval strategies more frequently in the printed database environment than EO participants did in the electronic database environment. Since there was no significant difference between group means for measures of process effectiveness, this suggests that the two task environments were different in some respects.

Table 5

Analysis of Variance of Product Retrieval Effectiveness Ratio Scores by Group During Stage Two Introductory Activities

Source	df	MS	F	p
Treatment Group (A)	1	3.306	64.217	.0000
Error (S/A)	44	0.515		
Total	45			

Stage Two: Effectiveness and Efficiency of Information Retrieval

During Stage Two Event 3, measures of retrieval effectiveness and retrieval efficiency were collected to address research questions two and three.

Research question two asked, "Is there a significant difference between performance of the MTE group and the EO group on measures of retrieval effectiveness during subsequent information retrieval learning activities in a computerized environment?"

Using the process effectiveness standard, each participant's information retrieval effectiveness for the set of six information requests was examined. The obtained process effectiveness ratio scores ranged from 0.15 to 1.00, with a grand mean of 0.738. The process effectiveness mean ratio score for the MTE group was 0.032 lower than the EO group's process effectiveness mean ratio score. The descriptive statistics are reported in Table 6.

Table 6

Summary Statistics for Process Retrieval Effectiveness Ratio Scores During Stage Two Assessment Activities.

	n	Mean	Min	Max	SD
MTE group	23	0.722	0.15	1.00	0.221
EO group	23	0.754	0.41	0.97	0.167
Total sample	46	0.738	0.15	1.00	0.195

A one way analysis of variance of the process effectiveness ratio scores for the assessment activities did not reveal a significant difference in performance between the two groups, $F(1,44) = 0.302$, $p > .05$. While there was no significant difference in performance between the MTE and the EO treatment groups during either the introductory or assessment activities, both groups achieved higher process effectiveness mean scores on the assessment activities than they had on the introductory activities. A repeated measures analysis of variance was performed to determine if there were significant differences in between-set performance scores for the two groups. Preliminary univariate homogeneity of variance tests using Bartlett-Box tests yielded the following F-ratios: $F(1,5808) = .826$, $p = .363$, and $F(1,5808) = 1.650$, $p = .199$, indicating that the assumption of homogeneity of variance had not been violated. The results of the repeated measures ANOVA are summarized in Table 7. Examination of Table 7 reveals that the group by set interaction effect was non-significant, $F(1,44) = .39$, $p > .05$, as was the group main effect, $F(1,44) = .19$, $p > .05$. The main effect for set, however, was statistically significant, $F(1,44) = 4.40$, $p = .042$. Examination of relevant marginal means indicates that participants scored higher on the assessment items, as compared to the introductory items, regardless of treatment group.

Table 7

Analysis of Variance of Process Effectiveness Scores by Group With Repeated Measures on Set

Source	df	MS	F	p	
Between-Subjects					
Treatment Group(G)	1	.01	.19	.664	(ns)
S/G	44	.05			
Within-Subjects					
Set(T)	1	.03	4.40	.042	
GT	1	.00	.39	.537	(ns)
S/GT	44	.01			

Using the product effectiveness standard, participants' retrieval effectiveness scores for the information requests used during the assessment activities was also examined. The obtained product effectiveness ratio scores across MTE and EO groups ranged from 0.00 to 1.00, with a grand mean of

0.207. The MTE mean was 0.037 higher than the EO mean for product effectiveness. The descriptive statistics are reported in Table 8.

Table 8

Summary Statistics for Product Retrieval Effectiveness Ratio Scores During Stage Two Assessment Activities.

	n	Mean	Min	Max	SD
MTE group	23	0.225	0.00	1.00	0.278
EO group	23	0.188	0.00	0.83	0.263
Total sample	46	0.207	0.00	1.00	0.268

To determine if there was a significant difference in performance between groups, data were subjected to an analysis of variance. The main effect for group was not statistically significant, $F(1,45) = 0.206$, $p > .05$, indicating that there was no significant difference in product effectiveness performance between the MTE and the EO groups during the Stage Two assessment activities.

Research question three asked, "Is there a significant difference between the MTE group and the EO group on measures of retrieval efficiency during subsequent information retrieval learning activities in a computerized environment?"

Process information retrieval efficiency ratio scores ranged from 0.29 to 1.29. The mean efficiency score for the MTE group was .077 higher than the mean efficiency score for the EO group. These descriptive statistics are reported in Table 9.

Table 9

Summary Statistics for Process Information Retrieval Efficiency Scores During Stage Two Assessment Activities.

	n	Mean	Min	Max	SD
MTE group	23	0.771	0.29	1.29	0.284
EO group	23	0.694	0.35	1.05	0.198
Total sample	46	0.733	0.29	1.29	0.245

To determine if there was a significant difference in performance between groups, data were subjected to an analysis of variance. The main effect for group was not statistically significant, $F(1,44) = 1.131$, $p > .05$, indicating that there was no significant difference in process retrieval efficiency performance between the MTE and the EO groups during the Stage Two assessment activities.

Product information retrieval efficiency ratio scores ranged from 0.00 to 1.02. The mean efficiency score for the MTE group was .058 higher than the mean efficiency score for the EO group. These descriptive statistics are reported in Table 10.

Table 10

Summary Statistics for Product Information Retrieval Efficiency Scores During Stage Two Assessment Activities.

	n	Mean	Min	Max	SD
MTE group	23	0.239	0.00	1.02	0.305
EO group	23	0.181	0.00	0.90	0.257
Total sample	46	0.210	0.00	1.02	0.280

To determine if there was a significant difference in performance between groups, data were subjected to an analysis of variance. The main effect for group was not statistically significant, $F(1,44) = 0.482$, $p > .05$, indicating that there was no significant difference in product retrieval efficiency performance between the MTE and the EO groups during the Stage Two assessment activities.

Stage Three: Effectiveness of Information Evaluation

Research question four asked, "Is there a significant difference between the MTE group and the EO group on measures of information evaluation?"

The Stage Three activities required participants to make 10 decisions about the accuracy of six sets of retrieved information. Participant's ratio evaluation effectiveness scores ranged from 0.533 to 1.00 with a grand mean of 0.791. The evaluation effectiveness mean ratio score for the MTE group was 0.019 lower than the evaluation effectiveness mean score for the EO Group. The descriptive statistics associated with the ratio scores for evaluation effectiveness are reported in Table 11.

Table 11

Summary Statistics for Evaluation Effectiveness Scores During Stage Three Activities.

	n	Mean	Min	Max	SD
MTE group	23	0.782	0.583	1.00	0.113
EO group	23	0.801	0.533	1.00	0.108
Total sample	46	0.791	0.533	1.00	0.110

To determine if there was a significant difference in performance between groups, data were subjected to an analysis of variance. The main effect for group was not statistically significant, $F(1,44) = 0.334$, $p > .05$, indicating that there was no significant difference between the performance of the MTE group and the EO group in evaluating the accuracy of retrieved information during the Stage Three activities.

Post Hoc Data Analysis

Subsequent to the analysis of the four research questions, examination of the descriptive data for CRMT scores and information retrieval and evaluation performance measures suggested that additional analyses could be undertaken to provide a better understanding of the nature of the information retrieval task, and the relationship between the defined entry level skills for this study and subsequent performance on the retrieval tasks.

The first question addressed during post hoc analysis was, "What is the degree of relationship between level of mastery of the entry level skills and subsequent measures of information retrieval and evaluation performance?"

One prerequisite for a participant's results to be included in the planned data analysis for this study was that the participant had to achieve mastery of the entry level knowledge and skills specified in the systematically designed instructional package, "Database Concepts and Skills". Demonstration of mastery was defined as a score of 80% or higher on the CRMT. Despite this prerequisite, data were collected for all volunteers (n = 79) who undertook, and completed, the full set of information retrieval activities so that the volunteers could receive course credit for their participation. Table 12 summarizes the distribution of these participants across treatment groups and mastery levels.

Table 12

<u>Distribution of Participants Across Treatment Groups and Mastery Level</u>

	Treatment Group	
	MTE	EO
Mastery	23	23
Non-mastery	15	18

When examining the descriptive statistics for measures of performance for each of these four groups, it was apparent that differences between means across mastery levels were markedly larger than differences between means across treatment groups for retrieval effectiveness, retrieval efficiency, and evaluation effectiveness. Tables 13 through 15 display the means for each of these three variables across mastery level and treatment group.

Table 13

<u>Means and Standard Deviations of Information Retrieval Effectiveness Scores Across Mastery Levels and Treatment Groups</u>.

	Treatment Group	
Mastery Level	MTE	EO
Mastery :		
Mean	.722	.754
Std. Dev.	.221	.167
Non-mastery:		
Mean	.554	.604
Std. Dev.	.250	.149

Table 14

Means and Standard Deviations of Information Retrieval Efficiency Scores Across Mastery Levels and Treatment Groups.

	Treatment Group	
Mastery Level	MTE	EO
Mastery:		
Mean	.771	.694
Std. Dev.	.284	.198
Non-mastery:		
Mean	.601	.567
Std. Dev.	.217	.165

Table 15

Means and Standard Deviations of Information Evaluation Effectiveness Scores Across Mastery Levels and Treatment Groups.

	Treatment Group	
Mastery Level	MTE	EO
Mastery:		
Mean	.782	.801
Std. Dev.	.113	.108
Non-mastery:		
Mean	.672	.681
Std. Dev.	.125	.117

To determine if there was a significant difference in subsequent performance between those participants who demonstrated mastery of entry level skills on the CRMT and those who did not, data were subjected to analyses of variance for the measures of retrieval effectiveness, retrieval efficiency and evaluation effectiveness. The main effects for mastery level were statistically significant in each of these analyses, $F(1,77) = 21.525$, $p < .001$, $F(1,77) = 12.275$, $p < .001$, $F(1,77) = 8.719$, $p = .0042$. These results indicated that there was a significant difference in performance between the mastery and the non-mastery participants across treatment type for each of these three measures of performance during the Stage Two and Stage Three activities. The results of these analyses are summarized in Tables 16 to 18.

Table 16

Analysis of Variance of Process Effectiveness Ratio Scores by Mastery Level During Stage Two Introductory Activities

Source	df	MS	F	p
Mastery Level (L)	1	.5258	21.5247	.0000
Error (S/L)	77	.0244		
Total	78			

Table 17

Analysis of Variance of Process Effectiveness Ratio Scores by Mastery Level During Stage Two Assessment Activities

Source	df	MS	F	p
Mastery Level (L)	1	.4742	12.2750	.0008
Error (S/L)	77	.0386		
Total	78			

Table 18

Analysis of Variance of Process Efficiency Ratio Scores by Mastery Level During Stage Two Assessment Activities

Source	df	MS	F	p
Mastery Level (L)	1	.4344	8.7193	.0042
Error (S/L)	77	.0498		
Total	78			

To determine the degree of relationship between the level of mastery of entry level skills and the measures of performance after controlling for Grade Point Average (GPA), first-order partial correlations were generated using the PARTIAL CORR procedure in SPSS-X to control for the effect of GPA on measures of information retrieval performance. These first order partial correlations, which are reported in Table 19, suggest that there is a significant relationship between mastery of the defined entry level skills and subsequent measures of information retrieval performance whereas the relationship between GPA and retrieval performance is not significant.

Table 19

First Order Partial Correlations Between CRMT Scores and Subsequent Information Retrieval and Evaluation Performances

Performance Measures	$r_{12.3}$	p
Process Retrieval Effectiveness (Stage Two, Introduction)	.4777	.0000
Process Retrieval Effectiveness (Stage Two, Assessment)	.4446	.0000
Process Retrieval Efficiency (Stage Two, Assessment)	.4282	.0000
Evaluation Effectiveness (Stage Three)	.5335	.0000

$r_{12.3}$ = first order partial correlation between mastery and performance measure after controlling for GPA

The second question addressed during post hoc analysis was, "Is information retrieval effectiveness related to the number of decisions required to implement the strategy?"

Table 20 details the mean retrieval effectiveness score by group for each information request during the Stage Two assessment activities. These results demonstrate that information requests requiring only six decisions were less effectively implemented than requests requiring between 13 and 20 decisions. This suggests that there is no simple linear relationship between the number of decisions required in the information retrieval strategy and the effectiveness with which that strategy is implemented.

Table 20

Mean Group Process Retrieval Effectiveness Scores For Each Information Request During Stage Two Assessment Activities

Request #	Decisions Required	MTE Group	EO Group	All Mastery
1	6	0.62	0.62	0.62
2	6	0.59	0.70	0.64
3	13	0.84	0.84	0.84
4	15	0.75	0.78	0.76
5	18	0.74	0.79	0.76
6	20	0.68	0.71	0.70

The third question addressed during post hoc analysis was, "Does performance on component tasks during implementation of an information retrieval strategy vary between mastery level groups?"

Descriptive data on the mean effectiveness with which the mastery and nonmastery groups implemented individual categories of component tasks during information retrieval are reported in Table 21. Three component task categories are listed: sort, selection and specification of fields.

Table 21

Mean Accuracy Ratios for Performance on Information Retrieval Component Tasks

Group	Sort	Selection	Fields
Mastery	0.77	0.76	0.65
Nonmastery	0.67	0.57	0.54

These results indicate marked differences between the mastery and nonmastery participants' performances on the three component tasks and suggest that accurate specification of display field types is a more difficult task for beginning computer users than the specification of sort and selection criteria.

To determine if there was a significant difference in strategy component performance between mastery and nonmastery participants, data were subjected to analyses of variance for the strategy component measures of sort effectiveness, selection effectiveness and field specification effectiveness. The main effects for mastery level were statistically significant in each of these analyses, $F(1,77) = 4.025$, $p = .048$, $F(1,77) = 17.712$, $p < .001$, $F(1,77) = 5.642$, $p = .020$. These results indicated that there was a significant difference in performance between the mastery and the non-mastery participants across treatment type for each of these three measures of strategy component performance during the Stage

Two assessment activities. The results of these analyses are summarized in Tables 22 to 24.

Table 22

Analysis of Variance of Sort Effectiveness Ratio Scores by Mastery Level During Stage Two Assessment Activities

Source	df	MS	F	p
Mastery Level (L)	1	.5322	4.025	.0483
Error (S/L)	77	.1322		
Total	78			

Table 23

Analysis of Variance of Selection Effectiveness Ratio Scores by Mastery Level During Stage Two Assessment Activities

Source	df	MS	F	p
Mastery Level (L)	1	1.2389	17.7121	.0001
Error (S/L)	77	.0699		
Total	78			

Table 24

Analysis of Variance of Field Specification Effectiveness Ratio Scores by Mastery Level During Stage Two Assessment Activities

Source	df	MS	F	p
Mastery Level (L)	1	.8551	5.6421	.0200
Error (S/L)	77	.1516		
Total	78			

CHAPTER V

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

In today's society, information is stored, manipulated and retrieved electronically. There is the expectation that teachers and students will be competent in the electronic retrieval of information, and that teachers will develop and incorporate instructional activities involving electronic information retrieval within a wide range of disciplines, with a particular focus on the development of problem solving skills. These expectations create a dilemma. Data retrieval systems are widely available within our schools for use by teachers and students, and teachers are expected to use these tools to create and manage problem solving activities. However, researchers have found that most teachers are novice computer users who cannot effectively retrieve information from electronic systems, and are even fearful of trying.

The following question must therefore be answered: "How can teachers be efficiently trained so that they can reliably retrieve valid information that has been stored electronically?" This question was the focus of this study.

To resolve the dilemma, and to find ways of efficiently training teachers who are novice computer users in the electronic retrieval of information, the following two questions were addressed:

1. What concepts and skills are needed to gain expertise in electronic information retrieval?

2. How can teachers who are novice computer users be efficiently trained to use these concepts and skills?

Researchers studying computer-human interaction have found that in tasks involving computer programming, word processing, and the use of electronic calculators, concrete experiences in non-computerized systems provided users with a conceptual model. This model enhanced their understanding of the abstract, electronic system with which they were interacting. This suggested that concrete experiences in information retrieval from a printed database during information retrieval training may provide the following advantages:

1. Working in a familiar environment may eliminate or reduce the anxiety associated with the computer environment;

2. The experiences in manual information retrieval may provide participants with a conceptual model from which they could develop a mental model of the information retrieval process;

3. Manual information retrieval may provide participants with intrinsic feedback as retrieval strategy decisions are implemented, thus allowing them to refine their strategies during execution of the task; and

4. Extending experiences from a printed to a computerized database might optimize performance since an isomorphic relationship exists between the tasks required in each of the retrieval environments.

For these reasons, two instructional approaches were envisaged, one involved manual then electronic (MTE) information retrieval experiences while the other involved electronic retrieval experiences only (EO). The manual experiences required participants to manipulate a printed database file to retrieve information, while the electronic experiences required participants to manipulate a computerized database file.

The study procedures used these distinct instructional approaches to aid participants in the development of information retrieval strategies. The effects of these treatments on subsequent skills of retrieval effectiveness, retrieval efficiency and evaluation effectiveness were then compared in a computerized database environment.

Eighty-two upper division undergraduate Education majors who were novice computer users participated in this study, which consisted of three instructional stages. Stage One provided instruction for entry level skills, Stage Two provided instruction for the development of information retrieval strategies, and Stage Three provided instruction for evaluating the accuracy of information retrieved from a database.

1. Is there a significant difference between measures of manual retrieval effectiveness and measures of electronic retrieval effectiveness during introductory learning activities in retrieval strategy implementation?

2. Is there a significant difference between performance of the MTE group and the EO group on measures of retrieval effectiveness during subsequent information retrieval learning activities in a computerized environment?

3. Is there a significant difference between the MTE group and the EO group on measures of retrieval efficiency during subsequent information retrieval learning activities in a computerized environment?

4. Is there a significant difference between the MTE group and the EO group on measures of information evaluation?

Conclusions

Research Questions

Based on the findings of this study, the following conclusions can be stated in response to the proposed research questions.

1. Participants who retrieved information from a computerized database file were significantly less effective ($p < .01$) in generating the required information product than participants who retrieved the same information from a

printed database file during the Stage Two introductory retrieval activities. However, there was no significant difference in measures of process retrieval effectiveness between participants retrieving information from the computerized versus the printed database file.

This suggests that the two task environments were different in some respects. Plausible differences include: (a) MTE participants received intermediate feedback as they manually retrieved information from the printed database, whereas EO participants received no feedback until the task was completed, (b) MTE participants were able to alter their decisions and modify their strategy during implementation whereas EO participants could not alter decisions that had been entered into the computer, and (c) MTE participants were able to repeat their retrieval strategies before recording their answers whereas EO participants were only given one opportunity to implement each retrieval strategy.

2. The medium in which participants initially retrieved information during training had no significant effect on the effectiveness of the participants' subsequent information retrieval performance in an electronic environment.

3. The medium in which participants initially retrieved information during training had no significant effect on the subsequent efficiency of the participants' information retrieval performance in an electronic

environment.

4. The medium in which participants initially retrieved information during training had no significant effect on the effectiveness of the participants' subsequent information evaluation performance.

These conclusions do not support the findings of Bayman and Mayer (1983) or Foss, Rosson and Smith (1982) who found that training involving a conceptual model was superior to procedural training alone. However they may provide some support for the findings of Douglas (1983), who found that differences in system processes between the conceptual model and the abstract system being modelled result in the mapping of inappropriate characteristics from the conceptual model to the system. The possible differences in the retrieval environments detailed above may have resulted in the MTE group not being so aware as the EO group of the need to focus attention on the decision-making processes when planning and evaluating their retrieval strategies, thereby nullifying any advantage that may otherwise have been gained by exposure to the conceptual model. Other possible explanations for these findings are (a) participants in the EO group were advantaged by receiving greater exposure to onscreen feedback than participants in the MTE group, and this advantage matched any advantage gained by exposure to the conceptual model, and (b) participants in both the MTE group and the EO group may have previously developed equally group and the EO group may have previously developed equally effective mental models of information retrieval systems based on experiences during Stage One instruction or informal experiences in information handling prior to the study.

The range in retrieval efficiency performance during the Stage Two assessment activities was particularly noticeable. These scores ranged from 0.29 to 1.29, indicating that some novice database users electronically retrieved information as efficiently as experienced database users after only six hours of instruction during the instructional activities in Stage Two of this study. This finding is similar to Kahney's (1983) findings regarding novice programmers; he noted that just as there are obvious differences between expert and novice programmers, there is an equally important distinction between "talented" and "average" programmers. Kahney noted that talented novices have much more in common with experts than they do with average novices. This was demonstrated in the current study by participants who, having achieved mastery of the entry level skills, were able to function at the competence level, the third stage in the model of skill acquisition described by Dreyfus and Dreyfus (1984, 1986).

<u>Post Hoc Investigations</u>

Additional conclusions, based on post hoc investigations, include:

1. There are significant relationships ($p < .001$) between the mastery of the defined entry level skills and concepts and the subsequent effectiveness and efficiency with which information is retrieved and evaluated. The procedures and results of this study provide three forms of support for the use of the <u>Database Concepts and Skills</u> instructional package during Stage One instruction, and use of the CRMT to define mastery of the skills prerequisite for the development and application of information retrieval strategies during Stage Two instruction. Content validity has been supported by the verification by subject-matter specialists that the performance objectives, and the criterion-referenced test items measuring those performances, were appropriate for the stated instructional goal. Construct validity has been supported by the findings that mastery participants' performances on measures of retrieval effectiveness, retrieval efficiency and evaluation effectiveness were significantly higher than the performances of nonmastery participants. Concurrent validity has been supported by the high positive correlations between CRMT performance and subsequent measures of retrieval effectiveness, retrieval efficiency

and evaluation effectiveness reported during this study.

2. The effectiveness with which participants retrieve information in response to an information request does not appear to be directly related to the apparent complexity of the information request, as measured by the number of decisions required to retrieve the requested information.

Closer examination of the nature of the information retrieval tasks provided some unexpected results. Table 5 documents the mean retrieval effectiveness scores for both the MTE group and the EO group, by question within the Stage Two assessment activities. Information requests one and two were the least effectively implemented. Yet each of these items required the fewest number of decisions, that is six decisions. Conversely, information request three, which required thirteen decisions, was most effectively implemented. These findings support the claim made by Katzeff (1986) that information requests requiring more steps for implementation are not necessarily the most difficult for the information retriever to conceptualize and implement. As Table 19 demonstrates information request six required 20 decisions. Yet this activity was more effectively implemented than information requests one and two, even though this item required participants to (a) formulate four separate selection criteria, (b) link those selection criteria with three logical operators, (c) group two selection criteria within parentheses, and (d) specify four display fields.

3. Mastery participants were significantly more effective in implementing the strategy components of sorting, selecting and field specification than nonmastery participants.

These comparisons of effectiveness of performance both between information requests within Stage Two assessment activities, and between strategy components for the same information requests, suggest that the beginning database users' perception of the task may be very different from the perception of an experienced database user. This compares with the findings of Berger and Wilde (1984), Chi, Feltovich and Glaser (1981), and Weiser and Shertz (1983) who found that novices are more distracted by surface features of the problem than the underlying structure of the problem task.

Implications

The significant differences in performance between participants who did and did not pass the CRMT highlight the gains in effectiveness and efficiency that can occur when novice database users master the entry level skills for information retrieval activities and apply this knowledge in conjunction with an algo-heuristic model when specifying information retrieval strategies. The results of this study also indicate that there is no significant advantage, or disadvantage, in using either a printed or an electronic database file when beginning database users initially apply these skills to develop information retrieval strategies. These findings have the following implications for database instruction:

1. Mastery of the entry level skills defined in this study appears to have a greater effect on subsequent information retrieval performance than the environment in which information retrieval instruction occurs. As these entry level skills may be gained without the use of a computer, allocation of limited computer resources in schools could be scheduled for the consolidation and application of well defined retrieval strategies, rather than for the development of the prerequisite skills and knowledge, and introduction of those strategies.

2. Introductory experiences in manually retrieving information from a printed database file neither positively nor adversely affects subsequent performance in electronic information retrieval from a computerized database file. In terms of efficient use of school computing resources, this suggests that information retrieval strategies can be developed effectively using a printed database, and later applied to computerized databases when access to computers is available. However, as shown in this study, these retrieval strategy skills can be rapidly acquired, and timely application of these skills to the computerized

environment for problem-solving activities should be encouraged.

3. The combination of mastery of entry level skills and the application of an algo-heuristic model for information retrieval, developed for use in this study, has resulted in some study participants achieving retrieval efficiency measures equal to, or greater than, some experienced database users. If similar results were obtained by high school students, it could be an indication that many of the educational database activities marketed by software publishers may not be sufficiently challenging for some students, and this further emphasizes the need for teachers to develop appropriate curriculum activities involving information retrieval.

Limitations of the Study

1. Participants in the study were drawn from a limited population and the results can not be generalized to populations of different ages and backgrounds.

2. The database user-interface provided functional, logical and syntactical simplicity for study participants, who were novice computer users; however, the participants were not able to alter their implemented strategy once a decision had been altered, and no opportunity was given to repeat an activity if a mistake had been made.

3. The instructional approaches to training provided the opportunity for study participants to achieve mastery of the defined entry level skills. However, participants were given no demonstration of, or active instruction in, the composite algo-heuristic model of information retrieval.

4. Only twelve records were used in the database file, which may have resulted in the MTE group viewing the manual information retrieval from the printed database file as a simplistic task. A larger printed database file may have required them to focus attention more closely on sort and selection criteria, and may have provided a more structured conceptual model of an information retrieval system.

5. Participants may have been disadvantaged by their lack of knowledge of the database contents, particularly the EO group, whose only reference was one sample record.

6. Although participants were novice computer users, their experience with information handling at home and in school may have already provided a conceptual model of information retrieval systems equivalent to that of the printed database.

Suggestions For Further Research

These recommendations for further study are derived from both a recognition of the limitations of this study, and from new questions that arose during the study.

1. As the participants for this study were college students, replication of the study with high school age students and inservice teachers would assess the effectiveness of the approach over the full span of ages for potential database users.

2. The limited duration of the study did not allow for continued database use and measures of changes in effectiveness and efficiency over an extended period of time. Longitudinal studies would provide answers to questions such as, "how long does it take for participants to reach maximum effectiveness?" and, "is it realistic to aim for performance effectiveness scores of 1.0 from all participants who have achieved mastery of the entry level skills and concepts?".

3. The use of strategy guides and the database user-interface in this study imposed a structure on the participants when they were making and implementing information retrieval decisions. The purpose of the algo-heuristic model in this study was to provide a model for retrieval strategy development that teachers could apply to a wide variety of discrete information systems, and that could be used independently by database users. Further research could investigate the effectiveness of applying the model with special purpose educational database software, both as an imposed structure with active demonstration of

the model and as an independently applied model.

4. The onscreen feedback provided during this study prompted many participants to ask questions concerning the difference between their displayed set of retrieved information and the answer. As a protocol during this study, no assistance of this type was given, yet it became apparent to the researcher that participants varied greatly in the extent to which they accommodated the feedback. A clinical investigation, monitoring the use of feedback by individual participants, could provide suggestions for developing approaches to the evaluation of information.

5. A similar study, focusing on decision analysis, could identify the most commonly occurring strategy errors so that instruction and remediation can become more effective.

6. A clinical investigation could be completed to monitor the use of onscreen feedback during information retrieval and to measure its effect on retrieval strategy refinement.

7. Classification of problem-type has been a focus of research in the fields of mathematics, physics and computer programming. An investigation of the classification of information requests by beginning database users could explain why seemingly simple information requests are handled least effectively by beginning database users.

8. Further research could be conducted to identify more precisely those prerequisite skills or attributes which lead to subsequent successful database use.

LIST OF REFERENCES

Alesandrini, K. L. (1982, April). Imagery-eliciting strategies and meaningful learning. Journal of Mental Imagery, 6, 125-140.

Allwood, C. M. (1986). Novices on the computer. International Journal of Man-Machine Studies, 25, 633-658.

Anderson, J. R. (1982). Acquisition of cognitive skills. Psychological Review, 89, 396-406.

Bates, M. J. (1986). Subject access in online catalogs: A design model. Journal of the American Society for Information Science, 37(6), 357-376.

Bateson, A. G., Alexander, R. A. & Murphy, M. D. (1987). Cognitive processing differences between novice and expert computer programmers. International Journal of Man-Machine Studies, 26, 649-660.

Bayman, P. & Mayer, R. E. (1983). A diagnosis of beginning programmers' misconceptions of BASIC programming statements. Communications of the Association for Computing Machinery (ACM), 26, 677-679.

Berger, D. E. & Wilde, J. M. (1984, March). Solving algebra word problems. Paper presented at the Claremont Conference on Applied Cognitive Psychology, Claremont, CA. (ERIC Document Reproduction Service No. ED 242 550)

Blohm, P. J. & Wiebe, J. H. (1980, December). Effects of representation and organizational features of text on success in mathematical problem-solving. Paper presented at the Annual Meeting of the American Reading Conference, Sarasota, FL. (ERIC Document Reproduction Service No. ED 195 930)

Bolt, R. (1986). The role of the computer in the teaching of social studies. Alberta, Canada. (ERIC Document Reproduction Service No. ED 282 802)

Borgman, C. L. (1986). The user's mental model of an information retrieval system: an experiment on a prototype online catalog. International Journal of Man-Machine Studies, 24, 47-64.

Bracey, R. M. L. (1984). Computers in education: an English viewpoint. In R. C. Barquin & G. P. Mead (Eds.), Towards the information society: Selected papers from the Hong Kong Computer Conference 1983 (pp.97-101). Amsterdam: Elsevier Science Publishers.

Brooks, L. W. & Dansereau, D. F. (1987). Transfer of information: An instructional perspective. In S. M. Cormier & J. D. Hagman (Eds.), Transfer of learning: Contemporary research and applications. San Diego, CA: Academic Press.

Brosey, M. & Shneiderman, B. (1978). Two experimental comparisons of relational and hierarchical database models. International Journal of Man-Machine Studies, 10, 625-637.

Brown, M. A. (1987). The computer as a tool for information processing in geography. History and Social Science Teacher, 22(4), 197-202.

Bruner, J. S. (1966). Toward a theory of instruction. Cambridge, MA: Belknap Press of Harvard University Press.

Bruner, J. S. (1985). Models of the learner. Educational Researcher, 14(6), 5-8.

Calfee, R. (1985). Computer literacy and book literacy. Educational Researcher, 14(5), 8-13.

Callison, Daniel. (1986). Using Wilsearch with high school students: A pilot study. *****

Carroll, J. M. & Mack, R. L. (1983). Actively learning to use a word processor. In W. Cooper (Ed.), Cognitive aspects of skilled typewriting. New York: Springer-Verlag.

Carroll, J. M. & Mack, R. L. (1984). Learning to use a word processor: by doing, by thinking, and by knowing. In J. C. Thomas & M. Schneider (Eds.), Human factors in computers. Norwood: Ablex.

Carroll, J. M. & Mack, R. L. (1985). Metaphor, computing systems, and active learning. International Journal of Man-Machine Studies, 22, 39-57.

Chi, M. T. H., Feltovitch, P. J. & Glaser, R. (1980). Categorization and representation of physics problems by experts and novices. Cognitive Science, 5, 121-152.

Collis, B. (1988). Computers, curriculum and whole-class instruction: Issues and ideas. Belmont, CA: Wadsworth.

Cooke, N. J. & Schvanelveld, R. W. (1988). Effects of computer programming experience on network representations of abstract programming concepts. International Journal of Man-Machine Studies, 29, 407-427.

Coombs, M. J., Gibson, R. & Alty, J. L. (1982). Learning a first computer language: strategies for making sense. International Journal of Man-Machine Studies, 16, 449-486.

Craik, K. (1943). The nature of explanation. Cambridge: Cambridge University Press.

Croft, W. B. & Thompson, R. H. (1987). I^3R: A new approach to the design of document retrieval systems. Journal of the American Society for Information Science, 38(6), 389-404.

de Kleer, J. & Brown, J. S. (1982, August). Some issues on mechanistic mental models. Proceedings of the Fourth Annual Conference of the Cognitive Science Society (pp. 7-9). Ann Arbor, Michigan.

Degl'Innocenti, R. & Ferraris, M. (1988). Database as a tool for promoting research activities in the classroom: An example in teaching humanities. Computers and Education, 12(1), 157-162.

Derr, R. L. (1984). Questions: Definitions, structure and classification. Reference Quarterly, 24(2), 186-190.

Di Sessa, A. A. (1981, August). The role of experience in models of the physical world. In Proceedings of the Third Annual Conference of the Cognitive Science Society (pp. 101-102). Berkeley, California.

Dick, W. & Carey, L. (1985). The systematic design of instruction (2nd ed.). Genview, IL: Scott, Foresman and Company.

Douglas, S. A. (1983). Learning to text edit: Semantics in procedural skill acquisition. (Doctoral Dissertation, Stanford University, Stanford, CA, 1983). Dissertation Abstracts International, 44, 1515B.

Dreyfus, H. L. and Dreyfus, S. E. (1984). Putting computers in their proper place: Analysis versus intuition in the classroom. Teachers College Record, 85, 578-601.

Dreyfus, H. L. and Dreyfus, S. E. (1986). Mind over machine: The power of human intuition and expertise in the era of the computer. New York: Free Press.

Du Boulay, B. & O'Shea, T. (1981). Teaching novices programming. In M. J. Coombs & J. L. Alty (Eds.)., Computing skills and the user interface. New York: Academic Press.

Du Boulay, B, O'Shea, T. & Monk, J. (1981). The black box inside the glass box: Presenting computing concepts to novices. International Journal of Man-Machine Studies 14, 237-249.

Dunlap, W. P. & McKnight, M. B. (1980). Teaching strategies for solving word problems in math. Academic THerapy, 15, 431-442.

Eason, K. D. & Damodaran, L. (1981). The needs of the commercial user. In M. J. Coombs and J. L. Alty (Eds.), Computing skills and the user interface. New York: Academic Press.

Edmonds, E. A. (1981). Adaptive man-computer interfaces. In M. J. Coombs and J. L. Alty (Eds.), Computing skills and the user interface. New York: Academic Press.

Elkerton, J. (1985). A behavioral evaluation of command-selection aids for inexperienced computer users. (Doctoral Dissertation, Virginia Polytechnic Institute and State University, 1985). Dissertation Abstracts International, 47, 330B.

Eisenstadt, M., Laubsch, J. & Kahney, H. (1981, August). Creating pleasant programming environments for cognitive science students. In Proceedings of the Third Annual Conference of the Cognitive Science Society (pp. 148-149). Berkeley, California.

Feigenbaum, E. A. (1977). The art of artificial intelligence: Themes and studies of knowledge engineering. In Proceedings of the International Joint Conference on Artificial Intelligence, 1977 (pp. 1014-1029).

Fitts, P. M. & Posner, M. I. (1967). Human performance. New York: Brooks.

Foss, D., Rosson, M. B. & Smith, P. (1982). Reducing manual labor: An experimental analysis of learning aids for a text editor. Proceedings of Human Factors in Computer Systems Conference. National Bureau of Standards, Gaithersburg. Maryland.

Frenzel, L. E. (1987). Crash course in artificial intelligence and expert systems. Indianapolis, IN: Howard W. Sams.

Gagne, R. M. Learnable aspects of problem solving. Educational Psychologist, 15(2), 84-92.

Gaines, B. R. (1979). Logical foundations for database systems. International Journal of Man-Machine Studies, 11, 481-500.

Gaskins, I. (1988). Teachers as thinking coaches: Creating strategic learners and problem solvers. Journal of Reading, Writing and Learning Disabilities, 4(1), 35-48.

Gentner, D. (1981, August). Generative analogies of mental models. In Proceedings of the Third Annual Conference of the Cognitive Science Society (pp. 97-100). Berkeley, California.

Gick, M. L. & Holyoak, K. J. (1985). Analogical problem Solving. In A. M. Aitkenhead And J. M. Slack (Eds.), Issues in cognitive modelling. London: Lawrence Erlbaum Associates. pp. 279-306.

Glaser, R. (1987). Teaching expert novices. Educational Researcher, 16()*** check vol and no, 5.

Gray, W. D. & Orasanu, J. M. (1987). Transfer of cognitive skills. In S. M. Cormier & J. D. Hagman (Eds.), Transfer of learning: Contemporary research and applications. San Diego, CA: Academic Press.

Green, T. R. G., Payne, S. J. & van der Veer, G. C. (Eds.). (1983). The psychology of computer use. London: Academic Press.

Greeno, J. G., James, C. T., DaPolito, F., & Polson, P. G. (1978). Associative learning: A cognitive analysis. Englewood Cliffs, NJ: Prentice-Hall.

Halasz, F. G. (1984). Mental models and problem solving in using a calculator. (Doctoral Dissertation, Stanford University, Stanford, CA, 1984). Dissertation Abstracts International, 45, 1046B.

Halasz, F. G. & Moran, T. P. (1983, December). Mental models and problem solving using a calculator. In A. Janda (Ed.), Human Factors in Computing Systems: Conference Proceedings of the Association for Computing Machinery (ACM), Special Interest Group on Computer and Human Interaction (SIGCHI) and the Human Factors Society. Boston, MA.

Hannah, Larry. (1987). Teaching data base search strategies. Computing Teacher, 14(9), 16-23.

Harter, S. P. and Peters, A. R. (1985). Heuristics for online information retrieval: A typology and preliminary listing. Online Review, 9(5), 407-424.

Hedberg, J. G. & Perry, N. R. (1984, April). Teacher cognitive styles and selection of computer courseware. Paper presented to the Annual Meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 244 621)

Heller, J. I. & Reif, F. (1982, August). Generation of useful problem representations in a semantically rich domain: The example of physics. Proceedings of the Fourth Annual Conference of the Cognitive Science Society (pp. 77-78). Ann Arbor, Michigan.

Hildreth, C. R. (1987). Beyond boolean: Designing the next generation of online catalogs. Library Trends, 35(4), 647-667.

Humphreys, M. S. & Yuille, J. C. (19710. Errors as a function of noun concreteness. Unpublished manuscript.

Hunter, B. (1984). My students use computers: Computer literacy in the K-8 curriculum. Reston, VA: Reston Publishing.

Hunter, B. (1985). Problem solving with data bases. Computing Teacher, 12(8), 20-27.

Jagodzinski, A. P. (1983). A theoretical basis for the representation of on-line computer systems to naive users. International Journal of Man-Machine Studies, 18, 215-252.

James, E. B. (1981). The user interface: How we may compute. In M. J. Coombs and J. L. Alty (Eds.), Computing skills and the user interface. New York: Academic Press.

Johnson, P. E., Zualkernan, I. & Garber, S. (1987). Specification of expertise. International Journal of Man-Machine Studies, 26, 161-181.

Johnson-Laird, P. N. (1981, August). The form and function of mental models. In Proceedings of the ThirdAnnual Conference of the Cognitive Science Society (pp.103-104). Berkeley, California.

Johnson-Laird, P. N. (1985). Mental Models. In A. M. Aitkenhead And J. M. Slack (Eds.), Issues in cognitive modelling. London: Lawrence Erlbaum Associates. pp. 81-100.

Johnston, J. (1985). Information literacy: Academic skills for a new age. Ann Arbor, MI: Ann Arbor Institute for Social Research. (ERIC Document Reproduction Service No. ED 270 042)

Kahney, H. (1983). Problem solving by novice programmers. In T. R. G. Green, S. J. Payne, & G. C. van der Veer (Eds.). The psychology of computer use. London: Academic Press. pp. 121-141.

Kahney, H. & Eisenstadt, M. (1982, August). Programmers' mental models of their programming tasks: The interaction of real-world knowledge and programming knowledge. In Proceedings of the Fourth Annual Cognitive Science Society Conference (pp.143-145). Ann Arbor, Michigan.

Katzeff, C. (1986). Dealing with a database query language in a new situation. International Journal of Man-Machine Studies, 25, 1-17.

Katzeff, C. (1988). The effect of different conceptual models upon reasoning in a database query writing task. International Journal of Man-Machine Studies, 29, 37-62.

Kaufmann, G. (1980). Imagery, Language and Cognition: Toward a theory of symbolic activity in human problem-solving. Oslo: Universitetsforlaget.

Kibirige, H. M. (1988). Computer-assisted reference services: What the computer will not do. Reference Quarterly, 27(3), 377-383.

Kieras, D. E. & Bovair, S. (1984). The role of a mental model in learning to operate a device. Cognitive Science, 8, 255-273.

Kolodner, J. L. (1983). Towards an understanding of the role of experience in the evolution from novice to expert. International Journal of Man-Machine Studies, 19, 497-518.

Landa, L. N. (1983). The algo-heuristic theory of instruction. In C. M. Reigeluth (Ed.), Instructional-design theories and models: An overview of their current status. (pp. 166-211). Hillsdale, NJ: Lawrence Erlbaum Associates.

Landa, L. N. (1987). A fragment of a lesson based on the algo-heuristic theory of instruction. In C. M. Reigeluth (Ed.), Instructional theories in action: Lessons illustrating selected theories and models (pp. 113-159). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lane, H. U. (Ed.). (1985). The world almanac and book of facts: 1986. New York: Newspaper Enterprise Association.

Lengel, J. G. (1987). Thinking skills, social studies, and computers. Social Studies, 78(1), 13-16.

Lewis, N. S. H. (1987). A study of the effects of concrete experiences on the problem-solving ability of tenth-grade students. Dissertation Abstracts International, 47, 2983A. (University Microfilms No. 8626454)

Lindamood, G. E. (1984). The informationalisation of T. C. MITS (The Celebrated Man In The Street). In R. C. Barquin & G. P. Mead (Eds.), Towards the information society: Selected papers from the Hong Kong Computer Conference 1983 (pp.27-32). Amsterdam: Elsevier Science Publishers.

Lindsay, P. H. & Norman, D. A. (1972). Human information processing: An introduction to psychology. New York: Academic Press.

Little, T. (1986). Microcomputers in social studies education. Michigan Social Studies Journal, 1(1), 33-35.

Lockard, J., Abrams, P. D. & Many, W. A. (1987). Microcomputers for educators. Boston: Little, Brown & Company.

Logan, G. D. (1985). Skill and automaticity: Relations, implications, and future directions. Canadian Journal of Psychology, 39(2), 367-386.

Martin, W. (1984). Touring an informational wonderland. Classroom Computer Learning, 4(7), 52-60.

Marzola, E. S. (1987). Using manipulatives in math instruction. Journal of Reading, Writing and Learning Disabilities International, 3(1), 9-20.

Mayer, R. E. (1975). Forward transfer of different reading strategies evoked by testlike events in mathematics text. Journal of Educational Psychology, 67, 165-169.

Mayer, R. E. (1976). Some conditions of meaningfuol learning for computer programming: Advance organizers and subject control of frame order. Journal of Educational Psychology, 68, 143-150.

Mayer, R. E. (1977). Thinking and problem solving: An introduction to human cognition and learning. Glenview, IL: Scott, Foresman and Company.

Mayer, R. E., Stiehl, C. C. & Greeno, J. P. (1975). Acquisition of understanding and skill in relation to subjects' preparation and meaningfulness of instruction. Journal of Educational Psychology, 67, 331-350.

Montgomery, H. & Allwood, C. M. (1978). On the subjective representation of statistical problems. Scandinavian Journal of Educational Research, 22(3), 107-127.

Moran, T. P. (1981). The command language grammar: a representation for the user interface of interactive computer systems. International Journal of Man-Machine Studies, 15(1), 3-50.

Moray, N. (1978). The strategic control of information processing. In G. Underwood (Ed.), Strategies of information processing. London: Academic Press. pp. 301-328.

Muth, K. D. (1982). Cognitive demands that arithmetic word problems impose on children. (ERIC Document Reproduction Service No. ED 223 442)

Newell, A. & Simon, H. (1972). Human problem solving. Englewood Cliffs, NJ: Prentice-Hall.

Norman, D. A. (1981). The trouble with UNIX. Datamation, 27, 139-150.

Norman, D. A. (1983). Some observations on mental models. In D. Gentner & A. L. Stevens (Eds.), Mental models. Hillsdale, NJ: Lawrence Erlbaum Associates. pp. 7-14.

Oborne, D. J. (1985). Computers at work: A behavioural approach. Chichester: John Wiley & Sons.

O'Hanlon, N. (1988). The role of library research instruction in developing teachers' problem solving skills. Journal of Teacher Education, 39(6), 44-49.

Oldroyd, B. K. & Schroder, J. J. (1982). Study of strategies used in online searching: 2. Positional logic - an example of the importance of selecting the right Boolean operator. Online Review, 6(2), 127-133.

Olson, D. R. (1985). Computers as tools of the intellect. Educational Researcher 14(5), 5-8.

Paige, J. M. & Simon, H. A. (1966). Cognitive processes in solving algebra word problems. In Kleinmuntz, B. (Ed.), Problem solving: Research, method, and theory. New York: John Wiley & Sons. pp. 51-118.

Parisi, L. (1985). Computer databases: Applications for the social studies. Washington, DC: National Institute of Education. (ERIC Document Reproduction Service No. ED 264 167)

Paxton, A. L. & Turner, E. J. (1984). The application of human factors to the needs of the novice computer user. International Journal of Man-Machine Studies, 20, 137-156.

Polya, G. (1945). How to solve it. Princeton University: Princeton University Press.

Pon, K. (1984). Databasing in the elementary (and secondary) classroom. Computing Teacher, 12(3), 28-30.

Pruitt, E., and Dowling, K. (1987). Searching with students online ... How we taught it and what we learned. School Library Media Activities Monthly, 3(9), 29-32.

Radecki, T. (1983). Generalized boolean methods of information retrieval. International Journal of Man-Machine Studies, 18, 407-409.

Rawitsch, D. (1988). The computerized database: Not a simple solution. Computing Teacher, 15(4), 34-37.

Ray, H. N. (1985). A study of the effect of different data models on casual users performance in writing database queries. International Journal of Man-Machine Studies, 23, 249-262.

Resnick, L. B. (1984). Cognitive science as educational research: Why we need it now. In Improving education perspectives on educational research (pp. 36-41). Pittsburgh, PA: National Academy of Education.

Resnick, L. B. (1987). Education and learning to think. Washington, D.C.: National Academy Press.

Resnik, H. (1984). From social studies to social science. Learning 13(3), 32-42.

Richardson, A. (1969). Mental imagery. London: Routledge & Kegan Paul.

Rooze, G. E. (1983, November). Integrating computer software into social studies instruction. Paper presented at the Convention of the National Council for the Social Studies, San Francisco. (ERIC Document Reproduction Service No. ED 239 586)

Rooze, G. E. (1986, November). Strategies for teaching students to process information using databases. Paper presented at the Annual Meeting of the National Council for the Social Studies, New York. (ERIC Document Reproduction Service No. ED 275 617)

Rooze, G. E. (1987, November). Developing thinking using databases: What's really involved. Paper presented at the Annual Meeting of the National Council for the Social Studies, Dallas. (ERIC Document Reproduction Service No. ED 290 670)

Salton, G. (1979). Mathematics and information retrieval. Journal of Documentation, 35(1), 1-29.

Sartor, R. (1986). Computers in social studies: Yesterday and today. Michigan Social Studies Journal, 1(1), 43-44.

Schroder, J. J. (1983). Study of strategies used in online searching: 3. Query refining. Online Review, 7(3), 229-236.

Segal, S. J. (Ed.). (1971). Imagery: Current cognitive approaches. London: Academic Press.

Shaw, D. (1986). Nine sources of problems for online searchers. Online REview, 10(5), 295-306.

Sheehan, P. W. (Ed.). (1972). The function and nature of imagery. New York: Academic Press.

Shneiderman, B. (1980). Software psychology: Human factors in computer and information systems. Cambridge, MA: Winthrop.

Shoderbek, P. P., Schoderbek, C. G. & Kefalis, A. G. (1985). Management systems: Conceptual considerations (3rd ed.). Plano, TX: Business Publications.

Shortliffe, E. H. (1976). Computer-based medical consultations: MYCIN. New York: American Elsevier.

Shute, G. E. (1979). Accounting students and abstract reasoning. Sarasota,FL: American Accounting Association.

Sneeringer, J. (1978). User-interface design for text editing: A case study. Software-Practice and Experience, 8, 543-557.

Soergel, D. (1982). Structure and construction of numeric databases. Drexel Library Quarterly, 18(3-4), 135-146.

Spence, R. & Apperley, M. A. (1982). Data base navigation: An office environment for the professional. Behavior and Information Technology, 1, 43-54.

Sternberg, R. J. (1986). Intelligence applied: Understanding and increasing your intellectual skills. San Diego, CA: Harcourt Brace Jovanovich.

Tapsfield, A. (1985). Where the grass is greener. Microscope, 16, 11-14.

Thomas, R. C. (1981). The design of an adaptable terminal. In M. J. Coombs and J. L. Alty (Eds.), Computing skills and the user interface. New York: Academic Press.

Thompson, C., & Vaughan, L. (Eds.) (1986). Computers in the classroom: Experiences teaching with flexible tools. Chelmsford, MA: Northeast Regional Exchange, Inc. (ERIC Document Reproduction Service No. ED 268 013)

Travis, B. P. (1981). Error analysis in solving algebra word problems. (ERIC Document Reproduction Service No. ED 209 095)

Troutman, A. P. & White, J. A. (1988). The micro goes to school: Instructional applications of microcomputer technology. Pacfic Grove, CA: Brooks/Cole.

Waterman, D. A. (1986). A guide to expert systems. Reading, MA: Addison-Wesley.

Weaver, D. (1986). Database software for social studies. Portland, OR: Northwest Regional Educational Laboratory. (ERIC Document Reproduction Service No. ED 273 518)

Weiser, M. & Shertz, J. (1983). Program problem representation in novice and expert programmers. International Journal of Man-Machine Studies, 19, 391-398.

White, C. S. (1986). The impact of structured activities with a computer-based file-management program on selected information-processing skills. (Doctoral Dissertation, Indiana University, 1985). Dissertation Abstracts International, 47, 513A.

Wiedenbeck, S. (1985). Novice/expert differences in programming skills. International Journal of Man-Machine Studies, 23, 383-390.

Young, R. M. (1981). The machine inside the machine: users' models of pocket calculators. International Journal of Man-Machine Studies, 15, 51-85.

Young, R. M. (1983). Surrogates and mappings:two kinds of conceptual mapping for interactive devices. In D. Gentner & A. L. Stevens (Eds.), Mental models. Hillsdale, NJ: Lawrence Erlbaum Associates. pp. 35-52.

APPENDIXES

APPENDIX A

DATABASE CONCEPTS AND SKILLS INSTRUCTIONAL MATERIALS

DATABASE CONCEPTS AND SKILLS

SECTION 1: Basic Terminology and Concepts

The term "data" refers to the complete collection of facts stored in a data file.

"Information" is the term used to refer to only those facts that are needed to answer a question.

Obtaining information from data involves the selection of meaningful data from the complete set of data available. It may also involve arranging the selected data in an order that is more useful to the person who asked the question.

For example, if a telephone user asks, "What is the phone number for the XYZ Repair Company", only one phone number from the thousands available in the directory will be selected to answer this question.

If a teacher asks, "Who are the students in the 1014 Home Room group", the information would include the names of the students, probably arranged either in alphabetical order by last name, or by student number.

1. Which of the following statements is correct?

___ (a) Data and information are two different terms that have the same meaning.

___ (b) Data is selected information.

___ (c) Information is selected data.

___ (d) Information is data that has been arranged in a meaningful way to the user.

___ (e) Data is information which has been arranged in a meaningful way to the user.

___ (f) Data is information which has been selected and arranged in a meaningful way to the user.

___ (g) Information is data that has been selected and arranged in a meaningful way to the user.

Feedback:

Answers (c), (d) and (g) are all correct.

Answer (g) combines the statements in (c) and (d) to give a more comprehensive answer.

Consider the following set of data:

STATE	CAPITAL CITY	NICKNAME	POPULATION	POP. DENSITY
Wyoming	Cheyenne	Equality State	511,000	5.3
Texas	Austin	Lone Star State	15,989,000	61.0
Louisiana	Baton Rouge	Pelican State	4,462,000	99.3
Florida	Tallahassee	Sunshine State	10,976,000	202.7
Iowa	Des Moines	Hawkeye State	2,910,000	52.0
Missouri	Jefferson	Show Me State	5,008,000	72.6
Nebraska	Lincoln	Cornhusker State	1,606,000	21.0
Utah	Salt Lake City	Beehive State	1,652,000	20.1
Colorado	Denver	Centennial State	3,178,000	30.6
Hawaii	Honolulu	The Aloha State	1,039,000	161.7
New Mexico	Santa Fe	The Land of Enchantment	1,424,000	11.7
West Virginia	Charleston	The Mountain State	1,952,000	81.1
Kentucky	Frankfort	Bluegrass State	3,723,000	93.9
Wisconsin	Madison	Badger State	4,766,000	87.5
Tennessee	Nashville	Volunteer State	4,717,000	114.1

There are a number of questions that could be asked regarding this small set of data for which information could be provided.

Which of the following questions could be answered using the data from this data set?

___ (a) What is the capital of Iowa?

___ (b) Which of the fifty states in the U.S. is the largest in area?

___ (c) How did Missouri get the nickname of the "Show Me" state?

___ (d) Which of the fifteen states shown is the most densely populated?

Feedback:

If you selected **(a)** and **(d)** as being the two questions that could be answered using data from the given data set, you understand that information can only be derived from a data set that contains data appropriate to the question.

A set of facts about a **single** object, place, person, event, etc., is called a **RECORD.**

For example, each entry in the telephone book represents one record; each record contains data for the subscriber's name, address, and telephone number.

Most of the data for that record was probably collected from a printed form that the subscriber completed when applying for the telephone service; of course the telephone company would have allocated the phone number and entered that data onto the form.

A collection of similar records (such as the set of records in a telephone directory) is called a **FILE,** and a collection of data files is referred to as a **DATABASE.**

Indicate whether the following statements are True or False by circling the letter **T** or **F.**

T F (a) A record is a collection of files.

T F (b) A single card or catalog entry for a book in a library catalog is an example of a record.

T F (c) The data representing student registration details for all students at the university is an example of a record.

T F (d) A telephone directory is an example of a file.

T F (e) The data that you enter onto a university registration form constitutes the data in one record.

T F (f) A file is a collection of databases.

Feedback:

If your answers were (a) F, (b) T, (c) F, (d) T, (e) T and (f) F congratulate yourself! As you can see from the answers, a record refers to only one object, person, etc., while a file is a collection of records.

When the data from a data file is printed, it is often displayed in the form of a table consisting of rows and columns. Consider the following set of data:

	Column 1 STATE	Column 2 FLOWER	Column 3 TREE	Column 4 BIRD
Row 1	Wyoming	Indian Paint Brush	Cottonwood	Meadowlark
2	Texas	Bluebonnet	Pecan	Mockingbird
3	Louisiana	Magnolia	Cypress	Eastern Brown Pelican
4	Florida	Orange Blossom	Sabal Palm	Mockingbird
5	Iowa	Wild Rose	Oak	Eastern Goldfinch
6	Missouri	Hawthorn	Dogwood	Bluebird
7	Nebraska	Goldenrod	Cottonwood	Western Meadowlark
8	Utah	Sego Lily	Blue Spruce	Seagull
9	Colorado	Rocky Mntn. Columbine	Colorado Blue Spruce	Lark Bunting
10	Hawaii	Hibiscus	Candlenut	Hawaiian Goose
11	New Mexico	Yucca	Pinon	Roadrunner
12	West Virginia	Big Rhododendron	Sugar Maple	Cardinal
13	Kentucky	Goldenrod	Kentucky Coffee Tree	Cardinal
14	Wisconsin	Wood Violet	Sugar Maple	Robin
15	Tennessee	Iris Tulip	Poplar	Mockingbird

In any data set that is displayed in table format, each row of the data table (other than the headings) represents a RECORD.

Each column of data represents a FIELD or type of data item.

Each column heading is the name of a field, or the FIELD NAME. The field name is used when we refer to specific types of data in the records of the data file.

In this data set there are fifteen records, each containing data for four fields. The four fields are: State, Flower, Tree, and Bird.

Circle the term that correctly completes each of the following statements:

(a) Row 7 of the data set is the (record, field) for Nebraska.

(b) Column 4 contains data for a single (record, field).

(c) The words Pecan, Oak, and Blue Spruce are data in the TREE (record, field).

(d) The set of data about Hawaii is an example of a (record, field).

(e) The field name is the (row, column) heading.

Feedback:

(a) record (b) field (c) field (d) record (e) column

Earlier, the term **data** was used to refer to the collection of facts stored in a data file. When we need to refer to the fact(s) stored in a single field of a single record, we use the terms "field contents", or "contents of a field".

Consider the following set of data:

STATE	POPULATION	RANK POP.	PER CAPITA INCOME	UNEMP RATE
Wyoming	511,000	49	$12,586	6.3%
Texas	15,989,000	3	$12,636	5.9%
Louisiana	4,462,000	18	$10,850	10.0%
Florida	10,976,000	6	$12,553	6.3%
Iowa	2,910,000	29	$12,090	7.0%
Missouri	5,008,000	15	$12,129	7.2%
Nebraska	1,606,000	36	$12,280	4.4%
Utah	1,652,000	35	$9,719	6.5%
Colorado	3,178,000	28	$13,742	5.6%
Hawaii	1,039,000	39	$12,761	5.6%
New Mexico	1,424,000	37	$10,330	7.5%
West Virginia	1,952,000	34	$9,846	15.0%
Kentucky	3,723,000	23	$10,374	9.3%
Wisconsin	4,766,000	16	$12,309	7.3%
Tennessee	4,717,000	17	$10,400	8.6%

Write down, in the space provided, the following data:

(a) The field contents for the UNEMP RATE field for Louisiana.

____

(a) The contents of the field for the POPULATION field for West Virginia.

____

Feedback:

(a) 10.0% (b) 1,952,000

SECTION 2: Selection Criteria

Deriving **information** from a data file often involves selecting records from the available data. For a record to be selected to become part of the information, the record must satisfy a **condition**, or set of **conditions**, in the information request.

One dictionary defines a condition as "a prerequisite" or "a modifying circumstance". In an information request, the condition is the word or phrase that requires us to determine whether or not a record will be selected for inclusion in the information list.

For example, if someone asked, "Which state (or states) have the mockingbird as the state bird", the phrase that tells us that a record must meet a prerequisite condition has been underlined.

Consider the following set of data:

STATE	CAPITAL CITY	NICKNAME	POPULATION	POP. DENSITY	AREA (sq.mi.)
Texas	Austin	Lone Star State	15,989,000	61.0	267,338
Louisiana	Baton Rouge	Pelican State	4,462,000	99.3	48,523
Iowa	Des Moines	Hawkeye State	2,910,000	52.0	56,290
Nebraska	Lincoln	Cornhusker State	1,606,000	21.0	77,227
Utah	Salt Lake City	Beehive State	1,652,000	20.1	84,916
Colorado	Denver	Centennial State	3,178,000	30.6	104,247
Hawaii	Honolulu	The Aloha State	1,039,000	161.7	6,450
West Virginia	Charleston	The Mountain State	1,952,000	81.1	24,181
Kentucky	Frankfort	Bluegrass State	3,723,000	93.9	40,395
Wisconsin	Madison	Badger State	4,766,000	87.5	56,154
Tennessee	Nashville	Volunteer State	4,717,000	114.1	42,244

For each of the following information requests, circle "yes" if the information request contains a condition, or circle "no" if no condition is included.

For those information requests that do include a condition, underline the word or phrase that expresses the condition.

(a) Which states have less than three million people? (yes no)

(b) Which state has Denver as its capital city? (yes no)

(c) Which state is known as the "Bluegrass State"? (yes no)

(d) Name the states larger than 100,000 sq. miles. (yes no)

(e) What are the names of each state's capital city? (yes no)

Feedback:

(a) yes: less than three million people

(b) yes: Denver as its capital city

(c) yes: known as the "Bluegrass State"

(d) yes: larger than 100,000 sq. miles

(e) no: the information produced as a result of this information request would contain data from every record in the data file. No condition would be used to select records from the data file.

Defining a Formal Selection Criterion

When retrieving information from a data file, we need to express the condition precisely. Doing this allows us to be sure that the records selected to produce the information are appropriate to the information request. One effective way of expressing each condition precisely is to set it out as a formal selection criterion. Each formal selection criterion consists of three parts:

(1) the field name of the field that we need to examine;
(2) a relational operator; and
(3) an argument.

We already know that the field name is the column heading used when the data is displayed in a table or rows and columns.

Relational operators are phrases or symbols that tell us how to compare the actual data we have with the type of data that we are looking for. Some phrases that act as relational operators (and their equivalent symbols, where relevant) include:

contains	
begins with	
is the same as	=
is equal to	=
matches	=
is less than	<
comes before	<
is greater than	>
comes after	>
is less than or equal to	<=
comes before or is the same as	<=
is greater than or equal to	>=
comes after or is the same as	>=
is not equal to	<>
does not match	<>

The specified argument completes the formal selection criterion. The simplest form of an argument is the value (either numeric or string/text/character) that you could expect to find as the field contents of the named field.

For example, the formal selection criteria for the information requests in the previous activity that included condition phrases are:

(a) Which states have less than three million people?

Selection Criterion: POPULATION < 3,000,000

(b) Which state has Denver as its capital city?

Selection Criterion: CAPITAL CITY = Denver

(c) Which state is known as the "Bluegrass State"?

Selection Criterion: NICKNAME = Bluegrass State

(d) Name the states larger than 100,000 sq. miles.

Selection Criterion: AREA (sq. mi.) > 100,000

(e) What are the names of each state's capital city?

Selection Criterion: None

Consider the following set of data:

STATE	CAPITAL CITY	NICKNAME	POPULATION
Texas	Austin	Lone Star State	15,989,000
Louisiana	Baton Rouge	Pelican State	4,462,000
Iowa	Des Moines	Hawkeye State	2,910,000
Nebraska	Lincoln	Cornhusker State	1,606,000
Utah	Salt Lake City	Beehive State	1,652,000
Colorado	Denver	Centennial State	3,178,000
Hawaii	Honolulu	The Aloha State	1,039,000
West Virginia	Charleston	The Mountain State	1,952,000
Kentucky	Frankfort	Bluegrass State	3,723,000
Wisconsin	Madison	Badger State	4,766,000
Tennessee	Nashville	Volunteer State	4,717,000

Generate the formal selection criterion for each of the following information requests:

(a) What is the nickname for the state of Texas?

__________ (field name) ________ (relational operator) ______________ (argument)

(b) What is the population of Iowa?

(field name)	(relational operator)	(argument)

(c) Which states have fewer than three million people?

(field name)	(relational operator)	(argument)

(d) Where is Nashville?

(field name)	(relational operator)	(argument)

Feedback:

(a) STATE = Texas; although the information being asked for is the nickname, the request specifies that the information is required only for the state of Texas.

(b) STATE = Iowa; although the information being asked for is the population, the request specifies that the information is required only for the state of Iowa.

(c) POPULATION < 3,000,000; although the information being asked for are the names of the states, the request specifies that the information is required only for those states that have a population that is less than three million.

(d) CAPITAL CITY = Nashville; although the information being asked for is "where" (the name of the state), the request specifies that the information is required only for the state that has the city of Nashville.

Combining Two Selection Criteria

Many information requests involve two conditions. After representing each condition as a formal selection criterion, we must then decide from the wording of the information request whether or not both, or only one, of the conditions must be satisfied for the record to be selected.

For example, an information request such as

> Which states with an unemployment rate greater than 7% have a population exceeding 5 million?

involves two conditions; the first is that the unemployment rate must be greater than 7%, while the second condition is that the population for that state must be greater than 5 million.

The way the information request is stated, both conditions would have to be true for a record to be selected. When representing the formal selection criteria for this information request, the logical operator AND is used to join the two criteria. This signifies that BOTH conditions must be satisfied. The formal selection criteria would be represented as:

UNEMP RATE > 7% AND POPULATION > 5,000,000

Another information request, such as

> Which states have a per capita income that is $13,000 or more, or an unemployment rate that is less than 5%?

also involves two conditions, but in this case only one of the two conditions must be satisfied for the record to be selected. The logical operator OR is used to join these two criteria. The formal selection criteria would be represented as:

PER CAPITA INCOME > $13,000 OR UNEMP RATE < 5%

Consider the following set of data:

STATE	JOINED UNION	PRINCIPAL CROPS	PER CAPITA INCOME	UNEMP RATE
Wyoming	07/10/1890	wheat, beans, barley, oats, sugar beets	$12,586	6.3%
Texas	12/29/1845	cotton, sorghum, grains, veg, citrus	$12,636	5.9%
Louisiana	04/30/1812	soybeans, sugarcane, rice, cotton	$10,850	10.0%
Florida	03/03/1845	citrus, vegetables, avocados, sugarcane	$12,553	6.3%
Iowa	12/28/1846	corn, soybeans, oats, hay	$12,090	7.0%
Missouri	08/10/1821	soybeans, corn, wheat, cotton	$12,129	7.2%
Nebraska	03/01/1867	corn, soybeans, hay, wheat, sorghum	$12,280	4.4%
Utah	01/04/1896	wheat, hay, apples, barley, alfalfa seed	$9,719	6.5%
Colorado	08/01/1876	corn, wheat, hay, sugar beets, barley	$13,742	5.6%
Hawaii	08/21/1959	sugar, pineapples, macadamia nuts	$12,761	5.6%
New Mexico	01/06/1912	wheat, hay, sorghum, grain, onions	$10,330	7.5%
West Virginia	06/20/1863	apples, peaches, hay, tobacco, corn	$9,846	15.0%
Kentucky	06/01/1792	tobacco, soybeans, corn, wheat, hay	$10,374	9.3%
Wisconsin	05/29/1848	corn, beans, beets, peas, hay, oats	$12,309	7.3%
Tennessee	06/01/1796	soybeans, tobacco, cotton, wheat, corn	$10,400	8.6%

Generate the selection criteria, including the logical operator (AND or OR) that formally describe the conditions expressed in the following information requests:

(a) I need a list of those states that joined the union before 1850 which have current average earnings per person greater than $12,000.

__________ (field name) ______ (relational operator) __________ (argument)

______ (logical operator)

__________ (field name) ______ (relational operator) __________ (argument)

(b) Which states have an unemployment rate that is in the 5% - 6% range?

(field name)	(relational operator)	(argument)

(logical operator)

(field name)	(relational operator)	(argument)

(c) How many states have tobacco or soybeans as one of their principal crops?

(field name)	(relational operator)	(argument)

(logical operator)

(field name)	(relational operator)	(argument)

(d) Which states that grow tobacco have an average personal income that is $11,500 or more?

(field name)	(relational operator)	(argument)

(logical operator)

(field name)	(relational operator)	(argument)

(e) Which states joined the union prior to, or after, the eighteen hundreds?

__________ (field name) __________ (relational operator) __________ (argument)

__________ (logical operator)

__________ (field name) __________ (relational operator) __________ (argument)

Feedback:

For these examples, there may be more than one way to accurately represent the formal selection criteria. Compare your answer with the range of possible alternatives provided. Notice that for these examples, where only two selection criteria are used, the order in which the criteria are written may be reversed.

(a) JOINED UNION < 01/01/1850 AND PER CAPITA INCOME > $12,000
PER CAPITA INCOME > $12,000 AND JOINED UNION < 01/01/1850
JOINED UNION <= 12/31/1849 AND PER CAPITA INCOME > $12,000
JOINED UNION < 01/01/1850 AND PER CAPITA INCOME >= $12,001

If you had recorded only the year part of the date in the criterion, remember that the argument should be stated in a way that "matches" the format of the original data.

(b) UNEMP RATE >= 5% AND UNEMP RATE <= 6%
UNEMP RATE < 6.1% AND UNEMP RATE > 4.9%
UNEMP RATE <= 6% AND UNEMP RATE >= 5%
UNEMP RATE >4.9% AND UNEMP RATE <= 6%

(c) PRINCIPAL CROPS contains tobacco OR PRINCIPAL CROPS contains soybeans
PRINCIPAL CROPS contains soybeans OR PRINCIPAL CROPS contains tobacco

If you had "=" as the relational operator instead of "contains", you hadn't taken into consideration that the PRINCIPAL CROPS field stores several facts, in any order, within the one field. All we can assume, if a state does have soybeans or tobacco as one of the principal crops, is that the word "soybeans" or "tobacco" will be contained somewhere within the field.

Knowing this, do you want to look at your answer for part (d) on the previous page before going on?

(d) PRINCIPAL CROPS contains tobacco AND PER CAPITA INCOME >= $11,500
PER CAPITA INCOME >= $11,500 AND PRINCIPAL CROPS contains tobacco
PRINCIPAL CROPS contains tobacco AND PER CAPITA INCOME > $11,499
PER CAPITA INCOME > $11,499 AND PRINCIPAL CROPS contains tobacco

(e) JOINED UNION < 01/01/1800 OR JOINED UNION > 01/01/1900
JOINED UNION > 01/01/1900 OR JOINED UNION < 01/01/1800
JOINED UNION <= 12/31/1799 OR JOINED UNION >= 01/01/1900
JOINED UNION >= 01/01/1900 OR JOINED UNION < 01/01/1800
JOINED UNION <= 12/31/1799 OR JOINED UNION > 12/31/1899
JOINED UNION > 12/31/1899 OR JOINED UNION <= 12/31/1799

Combining Three or More Selection Criteria

When an information request involves more than two conditions, the selection criteria must be defined, arranged, and linked by logical operators in such a way that only the records that meet the intended selection criteria are selected.

Parentheses can be used to group selection criteria linked by OR when all criteria in the group within parentheses must also be linked with another criterion, using the logical operator AND.

For example, if the letters A, B, C, and D represent four different selection criteria, we can predict which records would be selected by considering the order in which the pairs of criteria are listed, the use of parentheses to group criteria, and the logical operators used to link the criteria:

Example 1: A OR B OR C OR D

A record would be selected using these selection criteria if ANY ONE (or more) of the four conditions was true for that record.

Example 2: A AND B AND C AND D

A record would be selected using these selection criteria only if ALL of the four conditions were true for that record.

Example 3: A OR B OR C AND D

A record would be selected using these selection criteria if:

Both C and D were true; or
A was true; or
B was true.

Example 4: A OR B AND C OR D

A record would be selected using these selection criteria if:

Both B and C were true; or
A was true; or
D was true.

Example 5: A AND B OR C OR D

A record would be selected using these selection criteria if:

Both A and B were true; or
C was true; or
D was true.

Example 6: A AND B OR C AND D

A record would be selected using these selection criteria if:

A and B were true; or
C and D were true.

Example 7: A AND (B OR C) OR D

A record would be selected using these selection criteria if:

A was true, and either B *or* C was true; or
D was true.

Example 8: A AND (B OR C OR D)

A record would be selected using these selection criteria if:

A was true, and either B or C or D was true.

Example 9: A AND B AND (C OR D)

A record would be selected using these selection criteria if:

A was true and B was true and either C or D was true.

Example 10: A AND B AND C OR D

A record would be selected using these selection criteria if:

A and B and C were true; or
D was true.

Imagine that the letters W, X, Y and Z represent the following four selection criteria:

W: POPULATION > 3,000,000
X: LOCATION = West North Central
Y: PER CAPITA INCOME < $12,500
Z: PRINCIPAL CROPS contains oats

For each of the following examples, substitute the formal selection criterion for the letter, then determine which record(s), if any, from the following set of data meet the set of selection criteria.

	STATE	POPULATION	LOCATION	PRINCIPAL CROPS	PER CAPITA INCOME
1	Wyoming	511,000	Mountain	barley, oats, sugar beets	$12,586
2	Texas	15,989,000	Southwestern	grains, veg, citrus	$12,636
3	Louisiana	4,462,000	South Central Gulf	rice, cotton, melons	$10,850
4	Florida	10,976,000	Southeast Atlantic	avocados, sugarcane	$12,553
5	Iowa	2,910,000	Midwest	soybeans, oats, hay	$12,090
6	Missouri	5,008,000	West North Central	corn, wheat, cotton	$12,120
7	Nebraska	1,606,000	West North Central	hay, wheat, sorghum	$12,280
8	Utah	1,652,000	Middle Rocky Mntn	barley, alfalfa seed	$9,719
9	Colorado	3,178,000	West Central	sugar beets, barley	$13,742
10	Hawaii	1,039,000	Pacific	macadamia nuts, fruit	$12,761

Circle each record number for each record that would be selected using the following sets of criteria. If no records satisfy a set of criteria, circle (no records selected).

(a) W OR X OR Y OR Z

(1 2 3 4 5 6 7 8 9 10) (no records selected)

(b) W AND X AND Y AND Z

(1 2 3 4 5 6 7 8 9 10) (no records selected)

(c) W OR X OR Y AND Z

(1 2 3 4 5 6 7 8 9 10) (no records selected)

(d) W AND X OR Y OR Z

(1 2 3 4 5 6 7 8 9 10) (no records selected)

(e) W AND (X OR Y) OR Z

(1 2 3 4 5 6 7 8 9 10) (no records selected)

(f) W AND X AND (Y OR Z)

(1 2 3 4 5 6 7 8 9 10) (no records selected)

(g) W OR X AND Y OR Z

(1 2 3 4 5 6 7 8 9 10) (no records selected)

(h) W AND X OR Y AND Z

(1 2 3 4 5 6 7 8 9 10) (no records selected)

(i) W AND (X OR Y OR Z)

(1 2 3 4 5 6 7 8 9 10) (no records selected)

(j) W AND X AND Y OR Z

(1 2 3 4 5 6 7 8 9 10) (no records selected)

Feedback:

Did you select these records as being the ones that satisfied the stated criteria? Be sure to review the data and the selection criteria if your answers and these don't match.

(a) Each record <u>except</u> record 10.
(b) <u>No</u> records meet the stated criteria.
(c) 2, 3, 4, 5, 6, 7, 9
(d) 1, 3, 5, 6, 7, 8
(e) 1, 3, 5, 6
(f) Only record 6.
(g) 1, 2, 3, 4, 5, 6, 7, 9
(h) 5, 6
(i) 3, 6
(j) 1, 5, 6

<u>Generating Multiple Formal Selection Criteria.</u>

The following sequence can guide your decisions when generating multiple formal selection criteria:

1. Read the information request carefully to identify each condition.

2. Use the format FIELD NAME...*relational operator...argument* to formally specify a selection criterion for <u>each</u> condition.

3. Arrange the selection criteria in an order that allows you to connect pairs of criteria with appropriate logical operators (AND or OR).

4. Use parentheses to group selection criteria (usually those linked by OR) when each criterion in the group must also be linked with another criterion by the logical operator AND.

Consider the following set of data:

STATE	JOINED UNION	POPULATION	AREA (sq.mi.)	PRINCIPAL CROPS	UNEM RATE
Wyoming	07/10/1890	511,000	97,914	wheat, beans, barley, oats, sugar beets	6.3%
Texas	12/29/1845	15,989,000	267,338	cotton, sorghum, grains, veg, citrus	5.9%
Louisiana	04/30/1812	4,462,000	48,523	soybeans, sugarcane, rice, cotton, melons	10.0%
Florida	03/03/1845	10,976,000	58,560	citrus, vegetables, avocados, sugarcane	6.3%
Iowa	12/28/1846	2,910,000	56,290	corn, soybeans, oats, hay	7.0%
Missouri	08/10/1821	5,008,000	69,686	soybeans, corn, wheat, cotton	7.2%
Nebraska	03/01/1867	1,606,000	77,227	corn, soy beans, hay, wheat, sorghum	4.4%
Utah	01/04/1896	1,652,000	84,916	wheat, hay, apples, barley, alfalfa seed	6.5%
Colorado	08/01/1876	3,178,000	104,247	corn, wheat, hay, sugar beets, barley	5.6%
Hawaii	08/21/1959	1,039,000	6,450	sugar, pineapples, macadamia nuts, fruit	5.6%
New Mexico	01/06/1912	1,424,000	121,666	wheat, hay, sorghum, grain, onions	7.5%
West Virginia	06/20/1863	1,952,000	24,181	apples, peaches, hay, tobacco, corn	15.0%
Kentucky	06/01/1792	3,723,000	40,395	tobacco, soybeans, corn, wheat, hay	9.3%
Wisconsin	05/29/1848	4,766,000	56,154	corn, beans, beets, peas, hay, oats	7.3%
Tennessee	06/01/1796	4,717,000	42,244	soybeans, tobacco, cotton, wheat, corn	8.6%

Use the sequence described above to guide you as you generate formal selection criteria for the following information requests.

(a) What is the unemployment rate for those states that have a population between three and four million where wheat or barley is a principal crop?

(b) Are there any states which joined the union during the eighteen hundreds that have a population less than three million and an unemployment rate below 6%?

(c) Which states grow any of the following: corn, rice, oats or peas?

(d) Is there a state that is larger than 100,000 square miles, or has an unemployment rate between 7 and 9 percent, or more than four million people.

Feedback:

(a) Question: What is the unemployment rate for those states that have a population between three and four million where wheat or barley is a principal crop?

POPULATION >= 3,000,000 AND POPULATION <= 4,000,000 AND (PRINCIPAL CROP contains wheat OR PRINCIPAL CROP contains barley)

Providing the pair of selection criteria within parentheses remain together in parentheses, the order of these criteria may be reversed.

(b) Question: Are there any states which joined the union during the eighteen hundreds that have a population less than three million and an unemployment rate below 6%?

JOINED UNION >= 01/01/1800 AND JOINED UNION < 01/01/1900 AND POPULATION < 3,000,000 AND UNEMP RATE < 6%

These selection criteria may be arranged in any order because each pair of criteria are linked by the logical operator AND.

The criteria involving the JOINED UNION field may alternatively have been written as "JOINED UNION > 31/12/1799" and "JOINED UNION <= 31/12/1899"

(c) Question: Which states grow any of the following: corn, rice, oats or peas?

PRINCIPAL CROP contains corn OR PRINCIPAL CROP contains rice OR PRINCIPAL CROP contains oats OR PRINCIPAL CROP contains peas

These selection criteria may be arranged in any order because each pair of criteria are linked by the logical operator OR.

(d) Question: Which states are larger than 100,000 square miles, have an unemployment rate between 7 and 9 percent, or a population greater than four million people .

AREA (sq.mi.) > 100,000 OR UNEMP RATE >= 7% AND UNEMP RATE <= 9% OR POPULATION > 4,000,000

These selection criteria may be arranged in any order provided the two conditions linked by AND are kept together; the other two logical operators are OR.

SECTION 3: Determining Order

We can choose to arrange lists of data, consisting of words, numbers or dates, in order so that entries are easier to locate from a printed list.

For example:

Student records could be arranged in order by last name, or by student number.

Telephone subscriber information is listed in alphabetical order by the subscriber's last name in the standard directory.

Department stores could arrange printed lists of cancelled credit card numbers in ascending numerical order by card number.

Donors to charitable funds are often listed in descending numerical order by the size of their donation.

A travel agent might want a list of clients arranged in chronological order, according to their departure date.

The following rules usually apply when arranging data in ASCENDING order:

1. A space comes before a digit.
2. Smaller digits come before larger digits (0, 1, . . . 8, 9).
3. Digits come before letters of the alphabet.

These rules are reversed when arranging data in DESCENDING order.

When working with a data file, the field by which records are arranged, or SORTED, is called the KEY FIELD.

Examine each of the following sets of data. At the bottom of each data set, name the key field that has been used to SORT, or arrange the order of, the records. Circle the word *ascending* or *descending*, to indicate the order in which the records have been arranged by the key field.

STATE	POPULATION	POP. DENSITY	PER CAPITA INCOME	UNEMP RATE	BIRTHS PER YR	DEATHS PER YR
Texas	15,989,000	61.0	$12,636	5.9%	306,192	119,531
Florida	10,976,000	202.7	$12,553	6.3%	155,236	116,515
Missouri	5,008,000	72.6	$12,129	7.2%	78,517	52,332
Wisconsin	4,766,000	87.5	$12,309	7.3%	73,088	41,274
Louisiana	4,462,000	99.3	$10,850	10.0%	83,195	36,549
Kentucky	3,723,000	93.9	$10,374	9.3%	51,964	33,449
Colorado	3,178,000	30.6	$13,742	5.6%	54,471	20,941
Iowa	2,910,000	52.0	$12,090	7.0%	42,611	26,093
West Virginia	1,952,000	81.1	$9,846	15.0%	25,059	19,114
Utah	1,652,000	20.1	$9,719	6.5%	39,677	9,295
Nebraska	1,606,000	21.0	$12,280	4.4%	26,483	14,973
New Mexico	1,424,000	11.7	$10,330	7.5%	26,285	9,806
Hawaii	1,039,000	161.7	$12,761	5.6%	18,658	5,966
Wyoming	511,000	5.3	$12,586	6.3%	9,026	3,000

(a) Key field: ______________________ (ascending descending)

STATE	POPULATION	POP. DENSITY	PER CAPITA INCOME	UNEMP RATE	BIRTHS PER YR	DEATHS PER YR
Wyoming	511,000	5.3	$12,586	6.3%	9,026	3,000
Hawaii	1,039,000	161.7	$12,761	5.6%	18,658	5,966
West Virginia	1,952,000	81.1	$9,846	15.0%	25,059	19,114
New Mexico	1,424,000	11.7	$10,330	7.5%	26,285	9,806
Nebraska	1,606,000	21.0	$12,280	4.4%	26,483	14,973
Utah	1,652,000	20.1	$9,719	6.5%	39,677	9,295
Iowa	2,910,000	52.0	$12,090	7.0%	42,611	26,093
Kentucky	3,723,000	93.9	$10,374	9.3%	51,964	33,449
Colorado	3,178,000	30.6	$13,742	5.6%	54,471	20,941
Wisconsin	4,766,000	87.5	$12,309	7.3%	73,088	41,274
Missouri	5,008,000	72.6	$12,129	7.2%	78,517	52,332
Louisiana	4,462,000	99.3	$10,850	10.0%	83,195	36,549
Florida	10,976,000	202.7	$12,553	6.3%	155,236	116,515
Texas	15,989,000	61.0	$12,636	5.9%	306,192	119,531

(b) Key field: ______________________ (ascending descending)

STATE	POPULATION	POP. DENSITY	PER CAPITA INCOME	UNEMP RATE	BIRTHS PER YR	DEATHS PER YR
Colorado	3,178,000	30.6	$13,742	5.6%	54,471	20,941
Florida	10,976,000	202.7	$12,553	6.3%	155,236	116,516
Hawaii	1,039,000	161.7	$12,761	5.6%	18,658	5,966
Iowa	2,910,000	52.0	$12,090	7.0%	42,611	26,093
Kentucky	3,723,000	93.9	$10,374	9.3%	51,964	33,449
Louisiana	4,462,000	99.3	$10,850	10.0%	83,195	36,549
Missouri	5,008,000	72.6	$12,129	7.2%	78,517	52,332
Nebraska	1,606,000	21.0	$12,280	4.4%	26,483	14,973
New Mexico	1,424,000	11.7	$10,330	7.5%	26,285	9,806
Texas	15,989,000	61.0	$12,636	5.9%	306,192	119,531
Utah	1,652,000	20.1	$9,719	6.5%	39,677	9,295
West Virginia	1,952,000	81.1	$9,846	15.0%	25,059	19,114
Wisconsin	4,766,000	87.5	$12,309	7.3%	73,088	41,274
Wyoming	511,000	5.3	$12,586	6.3%	9,026	3,000

(c) Key field: ______________________ (ascending descending)

STATE	POPULATION	POP. DENSITY	PER CAPITA INCOME	UNEMP RATE	BIRTHS PER YR	DEATHS PER YR
Colorado	3,178,000	30.6	$13,742	5.6%	54,471	20,941
Hawaii	1,039,000	161.7	$12,761	5.6%	18,658	5,966
Texas	15,989,000	61.0	$12,636	5.9%	306,192	119,531
Wyoming	511,000	5.3	$12,586	6.3%	9,026	3,000
Florida	10,976,000	202.7	$12,553	6.3%	155,236	116,516
Wisconsin	4,766,000	87.5	$12,309	7.3%	73,088	41,274
Nebraska	1,606,000	21.0	$12,280	4.4%	26,483	14,973
Missouri	5,008,000	72.6	$12,129	7.2%	78,517	52,332
Iowa	2,910,000	52.0	$12,090	7.0%	42,611	26,093
Louisiana	4,462,000	99.3	$10,850	10.0%	83,195	36,549
Kentucky	3,723,000	93.9	$10,374	9.3%	51,964	33,449
New Mexico	1,424,000	11.7	$10,330	7.5%	26,285	9,806
West Virginia	1,952,000	81.1	$9,846	15.0%	25,059	19,114
Utah	1,652,000	20.1	$9,719	6.5%	39,677	9,295

(d) Key field: ______________________ (ascending descending)

Feedback:

(a) Population, descending
(b) Births Per Yr, ascending
(c) State, ascending
(d) Per Capita Income, descending

Fields containing dates are treated in a special way when records are being sorted. The date is considered to consist of three parts: the year, month and day portions of the field.

Dates in a date field will be sorted so that:

(a) all dates are first arranged in order by year;
(b) records having the same year will the be arranged by month;
(c) records having the same year and month will then be arranged by day of the month.

Examine each of the following sets of data. At the bottom of each data set, circle the option that best describes the order of data in the JOINED UNION field.

STATE	JOINED UNION	POPULATION	PER CAPITA INCOME	UNEMP RATE	TOURISM INCOME
Tennessee	06/01/1796	4,717,000	$10,400	8.0%	$2,100,000,000
Wisconsin	05/29/1848	4,766,000	$12,309	7.3%	$6,400,000,000
Kentucky	06/01/1792	3,723,000	$10,374	9.3%	$2,100,000,000
West Virginia	06/20/1863	1,952,000	$9,846	15.0%	$1,400,000,000
New Mexico	01/06/1912	1,424,000	$10,330	7.5%	$1,470,000,000
Hawaii	08/21/1959	1,039,000	$12,761	5.6%	$3,700,000,000
Colorado	08/01/1876	3,178,000	$13,742	5.6%	$4,000,000,000
Utah	01/04/1896	1,652,000	$9,719	6.5%	$754,000,000
Nebraska	03/01/1867	1,606,000	$12,280	4.4%	$1,200,000,000
Missouri	08/10/1821	5,008,000	$12,129	7.2%	$4,500,000,000
Iowa	12/28/1846	2,910,000	$12,090	7.0%	$1,600,000,000
Florida	03/03/1845	10,976,000	$12,553	6.3%	$20,000,000,000
Louisiana	04/30/1812	4,462,000	$10,850	10.0%	$3,300,000,000
Texas	12/29/1845	15,989,000	$12,636	5.9%	$125,000,000
Wyoming	07/10/1890	511,000	$12,586	6.3%	$750,000,000

(a) { sorted in ascending order by date, sorted in descending order by date, not sorted by date }

STATE	JOINED UNION	POPULATION	PER CAPITA INCOME	UNEMP RATE	TOURISM INCOME
Kentucky	06/01/1792	3,723,000	$10,374	9.3%	$2,100,000,000
Tennessee	06/01/1796	4,717,000	$10,400	8.6%	$2,100,000,000
Louisiana	04/30/1812	4,462,000	$10,850	10.0%	$3,300,000,000
Missouri	08/10/1821	5,008,000	$12,129	7.2%	$4,500,000,000
Florida	03/03/1845	10,976,000	$12,553	6.3%	$20,000,000,000
Texas	12/29/1845	15,989,000	$12,636	5.9%	$125,000,000
Iowa	12/28/1846	2,910,000	$12,090	7.0%	$1,600,000,000
Wisconsin	05/29/1848	4,766,000	$12,309	7.3%	$6,400,000,000
West Virginia	06/20/1863	1,952,000	$9,846	15.0%	$1,400,000,000
Nebraska	03/01/1867	1,606,000	$12,280	4.4%	$1,200,000,000
Colorado	08/01/1876	3,178,000	$13,742	5.6%	$4,000,000,000
Wyoming	07/10/1890	511,000	$12,586	6.3%	$750,000,000
Utah	01/04/1896	1,652,000	$9,719	6.5%	$754,000,000
New Mexico	01/06/1912	1,424,000	$10,330	7.5%	$1,470,000,000
Hawaii	08/21/1959	1,039,000	$12,761	5.6%	$3,700,000,000

(b) { sorted in ascending order by date, sorted in descending order by date, not sorted by date }

STATE	JOINED UNION	POPULATION	PER CAPITA INCOME	UNEMP RATE	TOURISM INCOME
Utah	01/04/1896	1,652,000	$9,719	6.5%	$754,000,000
New Mexico	01/06/1912	1,424,000	$10,330	7.5%	$1,470,000,000
Nebraska	03/01/1867	1,606,000	$12,280	4.4%	$1,200,000,000
Florida	03/03/1845	10,976,000	$12,553	6.3%	$20,000,000,000
Louisiana	04/30/1812	4,462,000	$10,850	10.0%	$3,300,000,000
Wisconsin	05/29/1848	4,766,000	$12,309	7.3%	$6,400,000,000
Kentucky	06/01/1792	3,723,000	$10,374	9.3%	$2,100,000,000
Tennessee	06/01/1796	4,717,000	$10,400	8.6%	$2,100,000,000
West Virginia	06/20/1863	1,952,000	$9,846	15.0%	$1,400,000,000
Wyoming	07/10/1890	511,000	$12,586	6.3%	$750,000,000
Colorado	08/01/1876	3,178,000	$13,742	5.6%	$4,000,000,000
Missouri	08/10/1821	5,008,000	$12,129	7.2%	$4,500,000,000
Hawaii	08/21/1959	1,039,000	$12,761	5.6%	$3,700,000,000
Iowa	12/28/1846	2,910,000	$12,090	7.0%	$1,600,000,000
Texas	12/29/1845	15,989,000	$12,636	5.9%	$125,000,000

(c) { sorted in ascending order by date, sorted in descending order by date, not sorted by date }

STATE	JOINED UNION	POPULATION	PER CAPITA INCOME	UNEMP RATE	TOURISM INCOME
Hawaii	08/21/1959	1,039,000	$12,761	5.6%	$3,700,000,000
New Mexico	01/06/1912	1,424,000	$10,330	7.5%	$1,470,000,000
Utah	01/04/1896	1,652,000	$9,719	6.5%	$754,000,000
Wyoming	07/10/1890	511,000	$12,586	6.3%	$750,000,000
Colorado	08/01/1876	3,178,000	$13,742	5.6%	$4,000,000,000
Nebraska	03/01/1867	1,606,000	$12,280	4.4%	$1,200,000,000
West Virginia	06/20/1863	1,952,000	$9,846	15.0%	$1,400,000,000
Wisconsin	05/29/1848	4,766,000	$12,309	7.3%	$6,400,000,000
Iowa	12/28/1846	2,910,000	$12,090	7.0%	$1,600,000,000
Texas	12/29/1845	15,989,000	$12,636	5.9%	$125,000,000
Florida	03/03/1845	10,976,000	$12,553	6.3%	$20,000,000,000
Missouri	08/10/1821	5,008,000	$12,129	7.2%	$4,500,000,000
Louisiana	04/30/1812	4,462,000	$10,850	10.0%	$3,300,000,000
Tennessee	06/01/1796	4,717,000	$10,400	8.6%	$2,100,000,000
Kentucky	06/01/1792	3,723,000	$10,374	9.3%	$2,100,000,000

(d) { sorted in ascending order by date, sorted in descending order by date, not sorted by date }

Feedback:

(a) Not sorted by date.
(b) Sorted in ascending order by date.
(c) Although, at first glance, this field appears to be in sorted order, the data has not been sorted by date.
(d) Sorted in descending order by date.

<u>When does information need to be arranged in order (or sorted)?</u>

Information needs to be arranged in order (sorted) before being presented to the user when:

1. The information request specifically asks for information to be arranged in a given order;

2. The information request includes terms such as first, last, most, least, highest, lowest, maximum, minimum, earliest, latest; these terms imply that data is arranged (or sorted) by a key field and that only the first (or last) items in the list will be useful in providing the necessary information to answer the question.

These two situations where data or information <u>must</u> be sorted could be described more economically as:

1. sort specified by user; and

2. sort for ranked presentation.

In addition to these two situations where the data or information <u>must</u> be sorted, there will be occasions when the retrieved information will be a list from which the user may need to refer to individual entries by a key field. For example, printed information about campus parking permits would be more useful if arranged by tag number, permit number, or student name, as these would probably be the three most common ways of referring to individual entries. Sometimes, conventional use of information requires a specific order. For example, class lists are usually arranged by student number or name.

Consider the following set of data:

STATE	JOINED UNION	TOURISM INCOME	MOTTO
Wyoming	07/10/1890	$750,000,000	Equal rights.
Texas	12/29/1845	$125,000,000	Friendship.
Louisiana	04/30/1812	$3,300,000,000	Union, justice and confidence.
Florida	03/03/1845	$20,000,000,000	In God we trust.
Iowa	12/28/1846	$1,600,000,000	Our liberties we prize and our rights we will maintain.
Missouri	08/10/1821	$4,500,000,000	The welfare of the people shall be the supreme law.
Nebraska	03/01/1867	$1,200,000,000	Equality before the law.
Utah	01/04/1896	$754,000,000	Industry.
Colorado	08/01/1876	$4,000,000,000	Nothing without providence.
Hawaii	08/21/1959	$3,700,000,000	The life of the land is perpetuated in righteousness.
New Mexico	01/06/1912	$1,470,000,000	It grows as it goes.
West Virginia	06/20/1863	$1,400,000,000	Mountaineers are always free.
Kentucky	06/01/1792	$2,100,000,000	United we stand, divided we fall.
Wisconsin	05/29/1848	$6,400,000,000	Forward.
Tennessee	06/01/1796	$2,100,000,000	Agriculture and commerce.

For each of the following information requests, circle the most appropriate sort specification. If a sort is required, enter the name of the key field.

(a) Which state, of the fifteen states listed, receives the most income from tourism?

(no sort required, sort specified by user, ranked sort)

Key field: ____________________

(b) I need a list showing each of the fifteen states and the date when each state joined the union.

(no sort required, sort specified by user, ranked sort)

Key field: ____________________

(c) Which state has the state motto "In God We Trust"?

(no sort required, sort specified by user, ranked sort)

Key field: ____________________

(d) Please provide an alphabetical list of the fifteen states showing each state's income from tourism.

(no sort required, sort specified by user, ranked sort)

Key field: ____________________

(e) People often ask me to tell them which state has a particular state motto. I need a list showing state names and the state motto.

(no sort required, sort specified by user, ranked sort)

Key field: ____________________

Feedback:

(a) ranked sort; key field: tourism income

(b) no sort required; however, you might choose to sort for convenience/convention; if so, key field: state

(c) no sort required; when only one record is expected to be selected when applying selection criteria, sorting is not necessary.

(d) sort specified by user; key field: state

(e) no sort required; however, you might choose to sort for convenience/convention; if so, key field: motto. If you had selected "state" as the key field, think about how the information list would be used. If you were asked to identify the state that has the motto "Mountaineers are always free" from the list, it would be more efficient to have the mottos arranged alphabetically than to work your way down a list arranged by state.

SECTION 4: Selecting Display Fields

Early in this instructional unit, the term "information" was described as selected data which may have been arranged in such a way that it can help the user answer a question.

In section two, you examined how formal selection criteria are developed to select the records which will be included in the "information". In this section, you will examine another important aspect of data selection during information retrieval. This aspect of selection limits the fields that will be displayed for each selected record.

When determining which fields should be displayed every field can be categorized as one, and only one, of the following:

(a) essential for display to provide the requested information;
(b) useful for display as a validating field; or
(c) unnecessary for display.

How do you decide that a field is essential for display?

Every information request will ask (either explicitly or implicitly) for specific information. The essential field(s) usually answer the question who, what, which, where or when?

Consider the following set of data:

STATE	POPULATION	TOURISM INCOME	UNEMP RATE	BIRTHS PER YR
Wyoming	511,000	$750,000,000	6.3%	9,026
Texas	15,989,000	$125,000,000	5.9%	306,192
Louisiana	4,462,000	$3,300,000,000	10.0%	83,195
Florida	10,976,000	$20,000,000,000	6.3%	155,236
Iowa	2,910,000	$1,600,000,000	7.0%	42,611
Missouri	5,008,000	$4,500,000,000	7.2%	78,517
Nebraska	1,606,000	$1,200,000,000	4.4%	26,483
Utah	1,652,000	$754,000,000	6.5%	39,677
Colorado	3,178,000	$4,000,000,000	5.6%	54,471
Hawaii	1,039,000	$3,700,000,000	5.6%	18,658
New Mexico	1,424,000	$1,470,000,000	7.5%	26,285
West Virginia	1,952,000	$1,400,000,000	15.0%	25,059
Kentucky	3,723,000	$2,100,000,000	9.3%	51,964
Wisconsin	4,766,000	$6,400,000,000	7.3%	73,088

For each of the following questions, circle only those fields that you consider <u>must be</u> displayed to provide information in response to the following information requests:

(a) Of the states listed, which has the largest population?

(state, population, tourism income, unemp rate, births per yr)

(b) What is the population of South Carolina?

(state, population, tourism income, unemp rate, births per yr)

(c) Which states have an unemployment rate greater than 8%?

(state, population, tourism income, unemp rate, births per yr)

(d) What income does New Mexico gain from tourism?

(state, population, tourism income, unemp rate, births per yr)

Feedback:

For each of those information requests, there was only one field that it would be essential to display. Did you select:

(a) state (b) population (c) state (d) tourism income ?

If you selected additional fields, you probably included those fields which are validating fields. Validating fields are those that include the data that verifies that the information being provided is the information that was requested.

Every field that is included in a formal selection criterion is a validating field.

For example, imagine that you were asked, "Which states earn more than five billion dollars in income from tourism?". The formal selection criterion would be TOURISM INCOME > $5,000,000,000. The only field that must be displayed would be the STATE field; however TOURISM INCOME is the validating field because it would show that the information from the retrieved records satisfies the stated criteria.

If the information request required that the selected information was sorted in either ascending or descending order so that the selection of a record or records could be made, the key field by which the data was sorted would also be a validating field for display purposes. For example, if we were asked to identify the state with the lowest unemployment rate, we could sort the records in descending order by unemployment rate, then select the first record. The only field that must be displayed would be the STATE field; however UNEMP RATE is the validating field because it would show that the actual value for the lowest unemployment rate from the retrieved record.

Using the same four questions, this time circle the field(s) that you could choose to display to validate the information selected in response to the information request:

(a) Of the states listed, which has the largest population?

(state, population, tourism income, unemp rate, births per yr)

(b) What is the population of South Carolina?

(state, population, tourism income, unemp rate, births per yr)

(c) Which states have an unemployment rate greater than 8%?

(state, population, tourism income, unemp rate, births per yr)

(d) What income does New Mexico gain from tourism?

(state, population, tourism income, unemp rate, births per yr)

Feedback:

For these information requests, the validating fields were:

(a) population (b) state (c) unemp rate (d) state.

Remember that the validating fields are either those fields which are specified in the selection criteria, or which are the key field by which the data is sorted, or both.

Now try some more examples. For each of the following questions, circle only those fields that you consider must be displayed to provide information in response to the following information requests:

(a) How much does Missouri earn from tourism each year?

{ state, population, tourism income, unemp rate, births per yr }

(b) What is the unemployment rate in Utah?

{ state, population, tourism income, unemp rate, births per yr }

(c) Which states have an unemployment rate less than 5%?

{ state, population, tourism income, unemp rate, births per yr }

(d) Where is the population the smallest?

{ state, population, tourism income, unemp rate, births per yr }

Using the same four questions, this time circle the field(s) that you could choose to display to validate the information selected in response to the information request:

(e) How much does Missouri earn from tourism each year?

{ state, population, tourism income, unemp rate, births per yr }

(f) What is the unemployment rate in Utah?

{ state, population, tourism income, unemp rate, births per yr }

(g) Which states have an unemployment rate less than 5%?

{ state, population, tourism income, unemp rate, births per yr }

(h) Where is the population the smallest?

{ state, population, tourism income, unemp rate, births per yr }

Feedback:

For each of those information requests, there was only one field that it would be essential to display. Did you select:

(a) tourism income (b) unemployment rate (c) state (d) state ?

For these information requests, the validating fields were:

(e) state (f) state (g) employment rate (h) population.

Now try some more examples. For each of the following questions, circle only those fields that you consider must be displayed to provide information in response to the following information requests:

(a) How much does Missouri earn from tourism each year?

{ state, population, tourism income, unemp rate, births per yr }

(b) What is the unemployment rate in Utah?

{ state, population, tourism income, unemp rate, births per yr }

(c) Which states have an unemployment rate less than 5%?

{ state, population, tourism income, unemp rate, births per yr }

(d) Where is the population the smallest?

{ state, population, tourism income, unemp rate, births per yr }

Using the same four questions, this time circle the field(s) that you could choose to display to validate the information selected in response to the information request:

(e) How much does Missouri earn from tourism each year?

{ state, population, tourism income, unemp rate, births per yr }

(f) What is the unemployment rate in Utah?

{ state, population, tourism income, unemp rate, births per yr }

(g) Which states have an unemployment rate less than 5%?

{ state, population, tourism income, unemp rate, births per yr }

(h) Where is the population the smallest?

{ state, population, tourism income, unemp rate, births per yr }

Feedback:

For each of those information requests, there was only one field that it would be essential to display. Did you select:

(a) tourism income (b) unemployment rate (c) state (d) state ?

For these information requests, the validating fields were:

(e) state (f) state (g) employment rate (h) population.

APPENDIX B

INSTRUCTIONAL GOAL, INSTRUCTIONAL ANALYSIS AND PERFORMANCE OBJECTIVES FOR THE DATABASE CONCEPTS AND SKILLS INSTRUCTIONAL MATERIALS

Instructional Goal:

Information retrieval procedures include identifying criteria by which records are selected from a data file, identifying the data items to be displayed in the output, and determining the order in which records are to be displayed. Given specific requests for information, and given details of the data items stored in a data file, the student will select the information retrieval procedures that would retrieve the required information from a printed set of data.

Instructional Analysis:

The four subordinate skills required by the learner before the instructional goal can be achieved are the ability to:

1. Identify the field names and the field contents for each field in the data file.

2. Generate selection criteria appropriate to the information request.

3. Select appropriate SORT option.

4. Select fields needed to display a validate the required information.

Each of these subskills is analyzed on the following pages.

Subskill 1: Identify the field names and the field contents for each field in the data file.

Recognize the difference between a field name and the contents of a field.	Define FIELD NAME or DATA ITEM
1.6	
Recognize the relationship between the terms "data" and "field contents".	Define CONTENTS OF A FIELD
1.5	
Recognize the difference between a record and a field.	Define FIELD Define RECORD
1.4	
Recognize the difference between a file and a record.	Define FILE
1.3	
Recognize the relationship between the terms "data" and "information".	
1.2	
Identify the difference between data and information.	Define INFORMATION Define DATA
1.1	

Subskill 2: Generate selection criteria appropriate to the information request.

Generate the set of formal criteria necessary to select records from a data file which satisfy a complex information request.
2.9

Identify the records which would be selected using multiple selection criteria linked by OR, AND and parentheses.
2.8

Define the effect of grouping selection criteria within parentheses.

Generate the formal criteria for an information request requiring two conditions linked by logical "OR".
2.7

Identify two criteria in an information request information request that need to be linked by logical "OR".
2.6

Define the meaning/effect of the logical connector "OR".

Generate the formal criteria for an information request requiring two conditions linked by logical "AND".
2.5

Identify two criteria in an information request that need to be linked by "AND".
2.4

Define the meaning/effect of the logical connector "AND".

Generate a single formal selection criterion for an information request.
2.3

Define the formal parts of a SELECTION CRITERION.

Recognize and use the symbols and/or phrases used to represent the range of relational operators. 2.2	Define the possible RELATIONAL OPERATORS.
Identify a single condition in an information request. 2.1	Define the term CONDITION.

Subskill 3: Select appropriate SORT option.

For a given information request, determine whether or not retrieved information needs to be sorted prior to display. If sorting is required, specify the key field, and the order in which records should be displayed. 3.5	
Recognize whether or not a set of data has been sorted in ASCENDING or DESCENDING order by a date field. 3.4	Describe how date fields are treated when they are used as key fields during a sort.
Recognize whether or not a set of data has been sorted in DESCENDING order. 3.3	Define DESCENDING ORDER.
Recognize whether or not a set of data has been sorted in ASCENDING order. 3.2	Define ASCENDING ORDER. Define KEY FIELD.
Identify sorted data from unsorted data. 3.1	Define SORT. Define conventional rules that govern order.

Subskill 4: Select fields needed to display a validate the required information.

Identify validating fields that may be displayed to validate the retrieved information. 4.2	Describe how to select the validating fields for a given information request. Define VALIDATING FIELD.
Identify essential fields that must be included in the set of displayed information. 4.1	Describe how to select the essential fields for a given information request. Define ESSENTIAL FIELD.

PERFORMANCE OBJECTIVES

1. Given a sample set of data from a data file, identify the field names and locate data for specified fields in specified records in the data file.

   1.1 Given statements concerning the relationship between DATA and INFORMATION, identify each statement as being true or false.

   1.2 Given actual data, or descriptions of data, classify the data as being either a RECORD or a FILE.

   1.3 Given actual data, or descriptions of data, classify a subset of the data as being either a RECORD or a FIELD.

   1.4 Given a data set, identify the data item that is the CONTENTS OF A FIELD for a specified field in a specified record.

   1.5 Given a data set, identify the field name for the field which contains specified field contents.

2. Generate selection criteria appropriate to a given information request.

   2.1 Given a written information request, identify and underline the phrase that specifies a single condition.

   2.2 Match the symbols used to represent the range of relational operators with the equivalent written phrase.

   2.3 Given a written information request, generate a single formal selection criterion that matches the condition in the information request.

   2.4 Given a written information request containing two conditions, underline the phrases which specify each condition and identify the logical operator--AND or OR--that is needed to link the two conditions.

   2.5 Given a written information request containing two conditions, generate the formal selection criteria that match the two conditions and indicate the logical operator--AND or OR--that is needed to link the two conditions.

2.6 Given a data set and a group of selection criteria, identify the records that would be selected using different groupings of multiple selection criteria linked by OR, AND and parentheses.

2.7 Generate the set of formal criteria necessary to select records from a data file that satisfy a complex information request.

3. Given an information request, select an appropriate SORT option.

3.1 Given a data set, determine whether or not data has been sorted by a key field.

3.2 Given a data set where the data has been sorted by a key field in ascending order, identify the key field.

3.3 Given a data set where the data has been sorted by a key field in descending order, identify the key field.

3.4 Given a data set containing a field where the data is in DATE format, determine whether or not the set of data has been sorted in ASCENDING or DESCENDING order by a date field.

3.5 For a given information request, determine whether or not retrieved information needs to be sorted prior to display. If sorting is required, specify the key field, and the order in which the records should be arranged.

4. Given an information request, select the fields which need to be displayed for output.

4.1 Given an information request and a data set, identify the essential fields which must be displayed to provide the required information.

4.2 Given an information request and a data set, identify the fields which, if displayed, would validate the information being provided.

APPENDIX C

CRITERION-REFERENCED MASTERY TEST FOR THE DATABASE CONCEPTS AND SKILLS INSTRUCTIONAL PACKAGE

CRITERION-REFERENCED MASTERY TEST

DATABASE CONCEPTS AND SKILLS

1. Select the term from the following list that best matches each of the descriptions given below. Place the letter corresponding to the matching term in the space provided.

_____(i) The term used to refer to selected data items that have been arranged in a way that is meaningful to the user is:

(a) selection criterion
(b) condition
(c) information
(d) key field

_____(ii) Facts about people, places, objects, events, etc., that have been collected and recorded are called:

(a) data
(b) condition
(c) information
(d) relational operator

_____(iii) The logical operator used to link two conditions so that a record will be selected during a search of a data file only if both of the two conditions are true is:

(a) or
(b) <>
(c) <=
(d) and

_____(iv) The phrase "is not equal to" can be represented by:

(a) or
(b) <>
(c) <=
(d) and

_____(v) A set of records which have the contents of a specified field listed showing the largest value first and the smallest value last is an example of a(n):

(a) a descending sort
(b) an ascending sort
(c) selection criteria
(d) logical operator

_____(vi) The value or set of characters recorded to represent a single item of data is called the:

(a) selection criterion
(b) field name
(c) key field
(d) contents of a field

_____(vii) The phrase "less than or equal to" can be represented by:

(a) <
(b) <=
(c) >=
(d) <>

_____(viii) A set of records of the same type are called a:

(a) record
(b) file
(c) field name
(d) condition

_____(ix) The part of a record which stores a single item of data is called the:

(a) key field
(b) field name
(c) file
(d) field

_____(x) The logical operator used to link two conditions so that a record will be selected during a search of a database if either, or both, of the two conditions are true is:

(a) parentheses
(b) and
(c) or
(d) <>

_____(xi) In a formal statement of a selection criterion, the symbol "=" means

(a) matches
(b) or
(c) and
(d) not equal to

_____(xii) The formal statement of the set of conditions necessary so that required records can be selected from the database is referred to as the:

(a) selection criteria
(b) field name
(c) contents of the field
(d) information

_____(xiii) A symbol which means "greater than" or "comes after" is:

(a) <
(b) <=
(c) >=
(d) >

_____(xiv) A symbol, or phrase, placed between the field name and the argument in a formal statement of a selection criterion is called a:

(a) logical operator
(b) relational operator
(c) condition
(d) primary key

_____(xv) The entire collection of facts are called:

(a) criteria
(b) information
(c) key fields
(d) data

_____(xvi) The field by which records within a file are arranged in order for display is the:

(a) field contents
(b) key field
(c) selection criterion
(d) condition

_____(xvii) A set of data relating to only one person, place, object, or event, is called a:

(a) condition
(b) field
(c) record
(d) file

_____(xviii) When data is arranged in ascending or descending order, it is said to be:

(a) selected
(b) relational
(c) logical
(d) sorted

_____(xix) A phrase that may be evaluated as being either "true" or "false" is called a:

(a) key field
(b) field name
(c) condition
(d) relational operator

_____(xx) The arrangement of data in alphabetical order or from smallest to largest values is an example of:

(a) a descending sort
(b) an ascending sort
(c) selection criterion
(d) logical operator

_____(xxi) The label used to identify a specific field in a database file is the:

(a) condition
(b) field contents
(c) file
(d) field name

2. Examine the small, but <u>complete</u>, set of data shown below. For convenience, the data have been arranged in numbered rows and columns, and each column has a heading.

	1	2	3	4	5	6	7	8
	SS_Number	Last Name	First Name	Destination	Flight #	Airline	Departure Date	Payment Date
1	721-54-3832	Seltmann	Rosa	Savannah	723	American	29-Jun-89	02-Apr-89
2	175-34-3933	Radney	Michelle	Miami	45	American	03-Jul-89	23-Mar-89
3	711-63-9126	Reiner	Jeanne	Austin	35	Continental	12-Jul-89	02-Apr-89
4	632-73-8521	Randazzo	Ann	Tulsa	312	American	15-Jul-89	12-Apr-89
5	318-53-2381	Regan	Leroy	Boulder	623	Northwest	18-Jul-89	03-Apr-89
6	826-25-7742	Boyer	Paul	Tacoma	63	Northwest	19-Jul-89	31-Mar-89
7	734-25-7127	Guida	Rosemary	Greensboro	52	Piedmont	01-Aug-89	03-Apr-89
8	471-43-2662	Katz	Shirley	Los Angeles	254	United	05-Aug-89	04-Apr-89
9	342-37-7142	Ammons	Phillip	Secaucus	272	United	10-Aug-89	01-Apr-89
10	729-11-6392	Baldwin	Richard	Reston	612	United	12-Aug-89	30-Mar-89
11	385-23-6523	Dykstra	Joel	Pittsburgh	902	Continental	22-Aug-89	02-Apr-89

Refer to the data set displayed above to complete the following statements.

For each statement, circle the term or phrase from the four alternatives provided that most correctly completes the statement.

a. The data in row 6 represents a (field, record, database, file).

b. All of the data shown above represents a (condition, field, record, file).

c. The word "Boulder" shown in (Row 5, Column 4) is an example of a (field name, field width, field contents, field type).

d. Other types of data sets, together with this data set, would constitute a (field, record, database, file).

e. Individual facts are stored as the (field name, field width, field contents, field type).

f. "DESTINATION" is the (field name, field width, field contents, field type) for column 4.

g. The data in column 7 have been sorted in (ascending, descending) order.

3. Indicate (by circling either the letter T or F) whether the following statements are true or false.

T F Data are collected by selecting information from a data file using selection criteria.

T F Information is usually more compact than the data from which it was obtained.

T F One form of information is the product obtained when data is arranged in ascending or descending order.

T F The data in a given database file will change with each information request, whereas the information produced will remain the same.

T F The data in a database file consists of the contents of every field for every record, whereas information may consist of the contents of selected fields for only those records that meet the selection criteria.

4. Imagine that the letters W, X, Y, and Z each represents a single selection criterion.

If only conditions X and Y were true for a particular record, would that record be selected from the data file if the selection criteria were:

(i)	W or X or Y or Z	(yes no)
(ii)	W or X or Y and Z	(yes no)
(iii)	W and X or Y and Z	(yes no)
(iv)	W or X or (Y or Z)	(yes no)
(v)	W and X and Y or Z	(yes no)
(vi)	W and (X or Y) and Z	(yes no)

4. When retrieving information from a database, the number of fields displayed is usually limited. Some fields must be displayed to provide the requested information. Other fields may be displayed to validate (or provide feedback regarding) the accuracy of the retrieved information in relation to the original information request.

Imagine a database file containing data about an airline booking system. The database file contains the following fields:

SS_Number	Passenger Name	Destination	Flight #
Airline	Departure Date	Payment Date	

For each of the following information requests, complete the following two steps.

a. Indicate (by circling the field names) those fields which are essential for providing the requested information.

(i) What is the destination for American Airline's flight 723?

SS_Number	Passenger Name	Destination	Flight #
Airline	Departure Date	Payment Date	

(ii) Which airline company has flights to Denver?

SS_Number	Passenger Name	Destination	Flight #
Airline	Departure Date	Payment Date	

(iii) When is Jane Dobkins leaving for New York?

SS_Number	Passenger Name	Destination	Flight #
Airline	Departure Date	Payment Date	

b. Indicate (by circling the field names) those fields which are validating fields.

(i) What is the destination for American Airline's flight 723?

SS_Number	Passenger Name	Destination	Flight #
Airline	Departure Date	Payment Date	

(ii) Which airline company has flights to Denver?

SS_Number	Passenger Name	Destination	Flight #
Airline	Departure Date	Payment Date	

(iii) When is Jane Dobkins leaving for New York?

SS_Number	Passenger Name	Destination	Flight #
Airline	Departure Date	Payment Date	

APPENDIX D

DATA SET FOR THE PRINTED AND ELECTRONIC DATABASE FILE

Record #: 1

COUNTRY: West Germany
CONTINENT: Europe
CAPITAL CITY: Bonn
CURRENCY: Mark
LANGUAGES SPOKEN: German
TYPE OF GOVERNMENT: Federal Republic
MAJOR INDUSTRIES: steel, ships, autos, machinery, coal, cement, chemicals
GROSS DOMESTIC PRODUCT: 583000000000
IMPORT COSTS: 153000000000
EXPORT EARNINGS: 171700000000
PER CAPITA INCOME: 9450
RADIOS (per 1,000 people): 396
TVs (per 1,000 people): 355
DAILY NEWSPAPER CIRCULATION (per 1,000 people): 141
MALE LIFE EXPECTANCY: 67.2
FEMALE LIFE EXPECTANCY: 73.4
LITERACY RATE: 99
CONSUMER PRICE INDEX (% annual change): 0.0
POPULATION: 61387000
POPULATION INCREASE (% annual change): 0.0

Record #: 2

COUNTRY: Brazil
CONTINENT: South America
CAPITAL CITY: Brasilia
CURRENCY: Cruzeiro
LANGUAGES SPOKEN: Portuguese,English
TYPE OF GOVERNMENT: Federal republic
MAJOR INDUSTRIES: steel, autos, chemicals, ships, appliances, shoes, paper
GROSS DOMESTIC PRODUCT: 192000000000
IMPORT COSTS: 15400000000
EXPORT EARNINGS: 26900000000
PER CAPITA INCOME: 1523
RADIOS (per 1,000 people): 633
TVs (per 1,000 people): 186
DAILY NEWSPAPER CIRCULATION (per 1,000 people): 44
MALE LIFE EXPECTANCY: 61.6
FEMALE LIFE EXPECTANCY: 65.7
LITERACY RATE: 75
CONSUMER PRICE INDEX (% annual change): 196.7
POPULATION: 134380000
POPULATION INCREASE (% annual change): 2.2

Record #: 3

COUNTRY: Argentina
CONTINENT: South America
CAPITAL CITY: Buenos Aires
CURRENCY: Austral
LANGUAGES SPOKEN: Spanish, English, Italian,German
TYPE OF GOVERNMENT: Military Republic
MAJOR INDUSTRIES: meat processing, flour milling, chemicals, textiles, machinery
GROSS DOMESTIC PRODUCT: 53000000000
IMPORT COSTS: 4500000000
EXPORT EARNINGS: 7800000000
PER CAPITA INCOME: 2331
RADIOS (per 1,000 people): 332
TVs (per 1,000 people): 196
DAILY NEWSPAPER CIRCULATION (per 1,000 people): 85
MALE LIFE EXPECTANCY: 66.8
FEMALE LIFE EXPECTANCY: 73.2
LITERACY RATE: 94
CONSUMER PRICE INDEX (% annual change): 626.7
POPULATION: 30097000
POPULATION INCREASE (% annual change): 1.5

Record #: 4

COUNTRY: Australia
CONTINENT: Australia
CAPITAL CITY: Canberra
CURRENCY: Dollar
LANGUAGES SPOKEN: English
TYPE OF GOVERNMENT: Parliamentary Democracy
MAJOR INDUSTRIES: iron, steel, textiles, autos, ships, aircraft, machinery, chemicals
GROSS DOMESTIC PRODUCT: 150200000000
IMPORT COSTS: 26400000000
EXPORT EARNINGS: 23700000000
PER CAPITA INCOME: 9960
RADIOS (per 1,000 people): 1293
TVs (per 1,000 people): 420
DAILY NEWSPAPER CIRCULATION (per 1,000 people): 426
MALE LIFE EXPECTANCY: 70.0
FEMALE LIFE EXPECTANCY: 76.0
LITERACY RATE: 100
CONSUMER PRICE INDEX (% annual change): 3.9
POPULATION: 15462000
POPULATION INCREASE (% annual change): 0.8

Record #: 5

COUNTRY: Afghanistan
CONTINENT: Asia
CAPITAL CITY: Kabul
CURRENCY: Afghani
LANGUAGES SPOKEN: Pashta, Dari Persian, Uzbek
TYPE OF GOVERNMENT: Dictatorship
MAJOR INDUSTRIES: textiles, carpets, cement
GROSS DOMESTIC PRODUCT: 3500000000
IMPORT COSTS: 562000000
EXPORT EARNINGS: 707000000
PER CAPITA INCOME: 168
RADIOS (per 1,000 people): 83
TVs (per 1,000 people): 1
DAILY NEWSPAPER CIRCULATION (per 1,000 people): 3
MALE LIFE EXPECTANCY: 39.9
FEMALE LIFE EXPECTANCY: 40.7
LITERACY RATE: 10
CONSUMER PRICE INDEX (% annual change): 4.9
POPULATION: 14448000
POPULATION INCREASE (% annual change): 2.2

Record #: 6

COUNTRY: United Kingdom
CONTINENT: Europe
CAPITAL CITY: London
CURRENCY: Pound
LANGUAGES SPOKEN: English, Welsh
TYPE OF GOVERNMENT: Constitutional monarchy
MAJOR INDUSTRIES: steel, metals, vehicles, shipbuilding, shipping, banking
GROSS DOMESTIC PRODUCT: 458000000000
IMPORT COSTS: 104800000000
EXPORT EARNINGS: 93700000000
PER CAPITA INCOME: 7216
RADIOS (per 1,000 people): 321
TVs (per 1,000 people): 334
DAILY NEWSPAPER CIRCULATION (per 1,000 people): 407
MALE LIFE EXPECTANCY: 70.2
FEMALE LIFE EXPECTANCY: 76.2
LITERACY RATE: 99
CONSUMER PRICE INDEX (% annual change): 5.0
POPULATION: 56023000
POPULATION INCREASE (% annual change): 0.0

Record #: 7

COUNTRY: Spain
CONTINENT: Europe
CAPITAL CITY: Madrid
CURRENCY: Peseta
LANGUAGES SPOKEN: Spanish,Catalan,Galician
TYPE OF GOVERNMENT: Constitutional Monarchy
MAJOR INDUSTRIES: machinery, steel, textiles, shoes, autos, ships
GROSS DOMESTIC PRODUCT: 160400000000
IMPORT COSTS: 28800000000
EXPORT EARNINGS: 23500000000
PER CAPITA INCOME: 5500
RADIOS (per 1,000 people): 271
TVs (per 1,000 people): 302
DAILY NEWSPAPER CIRCULATION (per 1,000 people): 116
MALE LIFE EXPECTANCY: 70.0
FEMALE LIFE EXPECTANCY: 76.0
LITERACY RATE: 97
CONSUMER PRICE INDEX (% annual change): 11.3
POPULATION: 38435000
POPULATION INCREASE (% annual change): 0.6

Record #: 8

COUNTRY: Canada
CONTINENT: North America
CAPITAL CITY: Ottawa
CURRENCY: Dollar
LANGUAGES SPOKEN: English, French
TYPE OF GOVERNMENT: Parliamentary Democracy
MAJOR INDUSTRIES: minerals, oil, meat products, fish, steel, electricity, agriculture
GROSS DOMESTIC PRODUCT: 317000000000
IMPORT COSTS: 78000000000
EXPORT EARNINGS: 90000000000
PER CAPITA INCOME: 10193
RADIOS (per 1,000 people): 1114
TVs (per 1,000 people): 493
DAILY NEWSPAPER CIRCULATION (per 1,000 people): 215
MALE LIFE EXPECTANCY: 69.0
FEMALE LIFE EXPECTANCY: 76.0
LITERACY RATE: 99
CONSUMER PRICE INDEX (% annual change): 4.3
POPULATION: 25142000
POPULATION INCREASE (% annual change): 0.7

Record #: 9

COUNTRY: Albania
CONTINENT: Europe
CAPITAL CITY: Tirana
CURRENCY: Lek
LANGUAGES SPOKEN: Albanian, Greek
TYPE OF GOVERNMENT: Socialist Republic
MAJOR INDUSTRIES: chemical fertilizers, textiles, electric cables
GROSS DOMESTIC PRODUCT: 2300000000
IMPORT COSTS: 250000000
EXPORT EARNINGS: 200000000
PER CAPITA INCOME: 830
RADIOS (per 1,000 people): 72
TVs (per 1,000 people): 7
DAILY NEWSPAPER CIRCULATION (per 1,000 people): 52
MALE LIFE EXPECTANCY: 69.0
FEMALE LIFE EXPECTANCY: 69.0
LITERACY RATE: 75
CONSUMER PRICE INDEX (% annual change): 0.0
POPULATION: 2906000
POPULATION INCREASE (% annual change): 2.0

Record #: 10

COUNTRY: Japan
CONTINENT: Asia
CAPITAL CITY: Tokyo
CURRENCY: Yen
LANGUAGES SPOKEN: Japanese
TYPE OF GOVERNMENT: Parliamentary democracy
MAJOR INDUSTRIES: electrical & electronic equip., autos, machinery, chemicals
GROSS DOMESTIC PRODUCT: 1200000000000
IMPORT COSTS: 136000000000
EXPORT EARNINGS: 149000000000
PER CAPITA INCOME: 8460
RADIOS (per 1,000 people): 776
TVs (per 1,000 people): 250
DAILY NEWSPAPER CIRCULATION (per 1,000 people): 569
MALE LIFE EXPECTANCY: 73.0
FEMALE LIFE EXPECTANCY: 78.0
LITERACY RATE: 99
CONSUMER PRICE INDEX (% annual change): 4.1
POPULATION: 119896000
POPULATION INCREASE (% annual change): 0.0

Record #: 11

COUNTRY: Libya
CONTINENT: Africa
CAPITAL CITY: Tripoli
CURRENCY: Dinar
LANGUAGES SPOKEN: Arabic
TYPE OF GOVERNMENT: Military Republic
MAJOR INDUSTRIES: carpets, textiles, shoes
GROSS DOMESTIC PRODUCT: 25000000000
IMPORT COSTS: 7300000000
EXPORT EARNINGS: 11100000000
PER CAPITA INCOME: 6335
RADIOS (per 1,000 people): 45
TVs (per 1,000 people): 46
DAILY NEWSPAPER CIRCULATION (per 1,000 people): 11
MALE LIFE EXPECTANCY: 56.1
FEMALE LIFE EXPECTANCY: 59.4
LITERACY RATE: 40
CONSUMER PRICE INDEX (% annual change): 0.0
POPULATION: 3684000
POPULATION INCREASE (% annual change): 3.5

Record #: 12

COUNTRY: United States
CONTINENT: North America
CAPITAL CITY: Washington,D.C.
CURRENCY: Dollar
LANGUAGES SPOKEN: English
TYPE OF GOVERNMENT: Republic
MAJOR INDUSTRIES: minerals, oil, meat products, fish, electricity, steel
GROSS DOMESTIC PRODUCT: 3701000000000
IMPORT COSTS: 341100000000
EXPORT EARNINGS: 217800000000
PER CAPITA INCOME: 11675
RADIOS (per 1,000 people): 2051
TVs (per 1,000 people): 605
DAILY NEWSPAPER CIRCULATION (per 1,000 people): 370 MALE LIFE EXPECTANCY: 71.6
FEMALE LIFE EXPECTANCY: 76.3
LITERACY RATE: 99
CONSUMER PRICE INDEX (% annual change): 4.3
POPULATION: 236413000
POPULATION INCREASE (% annual change): 0.6

APPENDIX E

SPECIFICATIONS FOR EACH SET OF INFORMATION REQUESTS

Type 1:

Question structure: No condition; Sort required

Type 2:

Question structure: One condition; No sort required

Type 3:

Question structure: Two conditions linked by OR; Sort required

Type 4:

Question structure: Three conditions linked by AND; No sort required

Type 5:

Question structure: Three conditions linked by AND, OR, & parentheses; Sort required

Type 6:

Question structure: Four conditions linked by AND, OR, & parentheses; No sort required

APPENDIX F

INFORMATION REQUESTS IN SETS A, B, C AND D

Set A:

1. Which country has the highest gross domestic product?

2. How many radios are there in Afghanistan for each one thousand people?

3. I need an alphabetical list of the countries where the cost of living increase is less than 5% or the average earnings are more than $9,000 a year.

4. Which non-Asian countries have a low number of radios (less than 300 for each one thousand people) but a reasonably high literacy rate (greater than 95%)?

5. I need an alphabetical list of the countries in Europe which have the pound or the mark as their unit of money?

6. Which non-European countries that have a small annual increase in their population (1.5% or less) also have English or French as one of their languages?

Set B:

1. Which country has the largest population?

2. Which country has the dinar as its unit of currency?

3. I need an alphabetical list of the names of the South American and Asian countries in this database file.

4. Which European countries have a consumer price index lower than 3% and an average income per person higher than $5,000?

5. I need an alphabetical list of the non-European countries where either English or German is spoken.

6. Are there any English speaking countries in Europe or South America that manufacture textiles?

Set C:

1. In which country do males have the shortest lifespan?
2. What is Brazil's unit of currency?
3. I need an alphabetical list of the countries in which the average life span for men or women is more than 70 years.
4. Which Asian countries have literacy levels between 95% and 99% (inclusive)?
5. Show me, from smallest to largest consumer price index increase, which English or Spanish speaking countries have annual consumer price index rises higher than three percent.
6. Which countries produce steel or manufacture machinery? (I don't want European or North American countries included in this list).

Set D:

1. Which country has the smallest population?
2. What is the per capita income of Argentina?
3. I need an alphabetical list of the countries where the either population is less than twenty million or the literacy level is greater than 98%.
4. Where could I go where the people speak English, have dollars as their unit of currency, and more than 95% of the population are literate?
5. Which Asian or African countries have more than 90% of the population literate? I'd like this information organized so that the countries with the higher literacy levels are shown first.
6. Are there any countries manufacturing textiles or fertilizer where the population is less than thirty million and the per capita income is more than five thousand dollars?

APPENDIX G

SAMPLE STRATEGY GUIDE

DECISION WORKSHEET - Set A

Please complete this worksheet as you work through the task of retrieving information from the printed database file.

When you have retrieved the information from the printed database file, record that information in the space provided below.

Information Request #2:

How many radios are there in Afghanistan for each one thousand people?

PART A: Sorting

Does the information request require that the database file, or the retrieved information, be sorted (arranged in order)?
Yes___ No___

If NO, move to Part B.

If YES, complete the next line:

Key sort field ______________ Order: Ascending___ Descending___

PART B: Selection Criteria

Must selection criteria be defined to retrieve the information? Yes___ No___

If NO, move to Part C on the following page.

If YES, how many separate selection criteria must be defined? 1___ 2___ 3___ 4___

In the space provided on the next page, enter each selection criterion by specifying the field name, relational operator, and argument.

When completed, if there are two or more selection criteria, enter the logical operator (AND or OR) needed to link them so that the required information will be retrieved.

Place parentheses around conditions which must be grouped.

____

Enter retrieved information here

______________ (field name) (relational operator) ______________ (argument)

(logical operator)

______________ (field name) (relational operator) ______________ (argument)

(logical operator)

______________ (field name) (relational operator) ______________ (argument)

(logical operator)

______________ (field name) (relational operator) ______________ (argument)

PART C: Essential Fields

Name the field(s) which must be displayed to provide the requested information:

__________ __________ __________ __________

PART D: Validating Fields

Name the field(s) which must be displayed to validate that the retrieved information is appropriate for this information request:

__________ __________ __________ __________

APPENDIX H

SAMPLE MENU SCREENS

Which field name needs to be specified in selection criterion1?

In the space provided below, type in the number of the field from the following list:

1	COUNTRY	11	PER CAPITA INCOME
2	CONTINENT	12	NUMBER OF RADIO SETS /1000
3	CAPITAL CITY	13	NUMBER OF TV SETS /1000
4	CURRENCY	14	DAILY NEWSPAPER CIRCULATION /1000
5	LANGUAGES	15	MALE LIFE EXPECTANCY
6	TYPE OF GOVERNMENT	16	FEMALE LIFE EXPECTANCY
7	MAJOR INDUSTRIES	17	LITERACY RATE
8	GROSS DOMESTIC PRODUCT	18	% CHANGE IN CONSUMER PRICE INDEX
9	VALUE OF EXPORTS	19	POPULATION
10	COST OF IMPORTS	20	% RATE OF POPULATION INCREASE

Enter the number for the field: 0

Select the relational operator for this selection criterion from the list below:

1	=	Equal to
2	<=	Less than or equal to
3	<	Less than
4	>	Greater than
5	>=	Greater than or equal to
6	<>	Not equal to
7	contains	

Type the number (1 - 7): 0

Enter the value or text in the space provided below for the argument that will complete this selection criterion.

COUNTRY =

Which field should be used as the key field (the field by which the data is arranged in order)?

Select the key field by typing the number of the field from the list below:

1 COUNTRY
2 CONTINENT
3 CAPITAL CITY
4 POPULATION
5 LANGUAGES
6 TYPE OF GOVERNMENT
7 MAJOR INDUSTRIES
8 GROSS DOMESTIC PRODUCT
9 VALUE OF EXPORTS
10 COST OF IMPORTS
11 PER CAPITA INCOME
12 NUMBER OF RADIO SETS
13 NUMBER OF TV SETS
14 DAILY NEWSPAPER CIRCULATION
15 MALE LIFE EXPECTANCY
16 FEMALE LIFE EXPECTANCY
17 LITERACY RATE
18 CONSUMER PRICE INDEX
19 CURRENCY
20 RATE OF POPULATION INCREASE

Enter the number corresponding to the key field: 0

Should the records be arranged in:

1 Ascending order
2 Descending order

Type 1 or 2 to select order of records: 0

APPENDIX I

CODING CRITERIA FOR SETS A - C

Each [*] represents a decision that needs to be made during information retrieval. When coding the data, one point was allocated for each correctly implemented decision.

Set A:

1. Which country has the highest gross domestic product?

   Sort: Yes [*], Gross Domestic Product [*], Descending [*]
   Conditions: None [*]
   Essential Fields: Country [*]
   Validating Fields: Gross Domestic Product [*]

   Possible Total: 6

2. How many radios are there in Afghanistan for each one thousand people?

   Sort: No [*]
   Conditions: Country [*] = [*] Afghanistan [*]
   Essential Fields: Radios [*]
   Validating Fields: Country [*]

   Possible Total: 6

3. I need an alphabetical list of the countries where the cost of living increase is less than 5% or the average earnings are more than $9,000 a year.

   Sort: Yes [*] Country [*] Ascending [*]
   Conditions: Cost Price Index [*] < [*] 5 [*] OR [*]
   Per Capita Income [*] > [*] 9000 [*]
   Essential Fields: Country [*]
   Validating Fields: Cost Price Index [*],
   Per Capita Income [*]

   Possible Total: 13

4. Which non-Asian countries have a low number of radios (less than 300 for each one thousand people) but a reasonably high literacy rate (greater than 95%)?

Sort: No [*]
Conditions: Continent [*] <> [*] Asia [*] AND [*]
Radios [*] < [*] 300 [*] AND [*]
Literacy Level [*] > [*] 95 [*]
Essential Fields: Country [*]
Validating Fields: Continent [*], Radios [*],
Literacy Level [*]

Possible Total: 16

5 I need an alphabetical list of the countries in Europe which have the pound or the mark as their unit of money?

Sort: Yes [*] Country [*] Ascending [*]
Conditions: Continent [*] = [*] Europe [*] AND [*]
([*] Currency [*] = [*] Mark [*] OR [*]
Currency [*] = [*] Pound [*])
Essential Fields: Country [*]
Validating Fields: Continent [*], Currency [*]

Possible Total: 18

6. Which non-European countries that have a small annual increase in their population (1.5% or less) also have English or French as one of their languages?

Sort: None [*]
Conditions: Continent [*] <> [*] Europe [*] AND [*]
Population Increase [*] <= [*] 1.5 [*] AND
[*] ([*] Languages [*] contains [*] English
[*] OR [*] Languages [*] contains [*] French
[*])
Essential Fields: Country [*]
Validating Fields: Continent [*], Population Increase
[*], Languages [*]

Possible Total: 21

Set B:

1. Which country has the largest population?

   Sort: Yes [*], Population [*], Descending [*]
   Conditions: None [*]
   Essential Fields: Country [*]
   Validating Fields: Population [*]

   Possible Total: 6

2. Which country has the dinar as its unit of currency?

   Sort: No [*]
   Conditions: Country [*] = [*] Argentina [*]
   Essential Fields: Per Capita Income [*]
   Validating Fields: Country [*]

   Possible Total: 6

3. I need an alphabetical list of the South American and Asian countries in this database file.

   Sort: Yes [*] Country [*] Ascending [*]
   Conditions: Population [*] < [*] 20000000 [*] OR [*]
   Literacy Level [*] > [*] 98 [*]
   Essential Fields: Country [*]
   Validating Fields: Population [*], Literacy Level [*]

   Possible Total: 13

4. Which European countries have a consumer price index lower than 3% and an average income per person higher than $5,000?

   Sort: No [*]
   Conditions: Languages[*] contains [*] English [*] AND [*]
   Currency [*] = [*] Dollar [*] AND [*]
   Literacy Level [*] > [*] 95 [*]
   Essential Fields: Country [*]
   Validating Fields: Languages [*], Currency [*],
   Literacy Level [*]

   Possible Total: 16

5. I need an alphabetical list of the non-European countries where either English or German is spoken.

   Sort: Yes [*] Literacy Level [*] Descending [*]
   Conditions: Literacy Level [*] > [*] 90 [*] AND [*]
   ([*] Continent [*] = [*] Asia [*] OR [*]
   Continent [*] = [*] Africa [*])
   Essential Fields: Country [*]
   Validating Fields: Literacy Level [*], Continent [*]

   Possible Total: 18

6. Are there any English speaking countries in Europe or South America that manufacture textiles?

   Sort: None [*]
   Conditions: Population [*] < [*] 30000000 [*] AND [*]
   Per Capita Income [*] > [*] 5000 [*] AND
   [*] ([*] Major Industries [*] contains [*]
   textiles [*] OR [*] Major Industries [*]
   contains [*] fertilizer [*])
   Essential Fields: Country [*]
   Validating Fields: Population [*], Per Capita Income [*],
   Major Industries [*]

   Possible Total: 21

Set C:

1. In which country do males have the shortest lifespan?

   Sort: Yes [*], Male Life Expectancy [*], Ascending [*]
   Conditions: None [*]
   Essential Fields: Country [*]
   Validating Fields: Male Life Expectancy [*]

   Possible Total: 6

2. What is Brazil's unit of currency?

   Sort: No [*]
   Conditions: Country [*] = [*] Brazil [*]
   Essential Fields: Currency [*]
   Validating Fields: Country [*]

   Possible Total: 6

3. I need an alphabetical list of the countries in which the average life span for men or women is more than 70 years.

   Sort: Yes [*] Country [*] Ascending [*]
   Conditions: Male Life Expectancy [*] > [*] 70 [*] OR [*]
   Female Life Expectancy [*] > [*] 70 [*]
   Essential Fields: Country [*]
   Validating Fields: Male Life Expectancy [*],
   Female Life Expectancy [*]

   Possible Total: 13

4. Which Asian countries have literacy levels between 95% and 99% (inclusive)?

Sort: No [*]
Conditions: Continent [*] = [*] Asia [*] AND [*]
Literacy Level [*] >= [*] 95 [*] AND [*]
Literacy Level [*] <= [*] 99 [*]
Essential Fields: Country [*]
Validating Fields: Continent [*], Literacy Level [*]

Possible Total: 15

5. Show me, from smallest to largest consumer price index increase, which English or Spanish speaking countries have annual consumer price index rises higher than three percent.

Sort: Yes [*] Consumer Price Index [*] Ascending [*]
Conditions:Consumer Price Index [*] < [*] 3 [*] AND [*]
([*] Languages [*] contains [*] English [*]
OR [*] Languages [*] contains [*] Spanish [*])

Essential Fields: Country [*]
Validating Fields: Consumer Price Index [*], Languages[*]
Possible Total: 18

6. Which countries produce steel or manufacture machinery? (I don't want European or North American countries included in this list).

Sort: None [*]
Conditions: Continent [*] <> [*] Europe [*] AND [*]
Continent [*] <> [*] North America [*] AND
[*] ([*] Major Industries [*] contains [*]
steel [*] OR [*] Major Industries [*]
contains [*] French [*])
Essential Fields: Country [*]
Validating Fields: Continent [*], Major Industries [*]

Possible Total: 20

APPENDIX J

SAMPLE EVALUATION CHECKLIST

Information Request #2: What is the per capita income of Argentina?

Information:

2331 Argentina

EVALUATION CHECKLIST

Instructions: Examine the information displayed above in conjunction with the printed data set and the information request. For each of the following questions, <u>circle</u> your response to each question.

1. Did the data (or retrieved information) need to be arranged in order? YES NO

   If **YES**,

   (a) Was the correct key field used? YES NO UNSURE

   (b) Was the data (or information) arranged in an appropriate order? YES NO UNSURE

2. Were one or more selection criteria required to select the appropriate information? YES NO

   If **NO**, were selection criteria used when they should not have been? YES NO UNSURE

   If **YES**, does it appear that all required selection criteria were used correctly to retrieve the required information? YES NO UNSURE

3. Were all <u>essential</u> fields displayed? YES NO UNSURE

   If **NO**, name the essential field (or fields) that should have been displayed.

   ____

4. Were all <u>validating</u> fields displayed? YES NO UNSURE

   If **NO**, name the validating field (or fields) that should have been displayed.

   ____

APPENDIX K

LISTING OF dBASE III PLUS COMMAND FILE USER-INTERFACE

```
clear
clear memory
use data
set index to
set status off
set talk off
PUBLIC setnum, qnum, date, starttime,midtime,endtime
PUBLIC totaltime,decidetime,entertime
store "A" to setnum
store 1 to qnum
store date() to date
do INTRO
do while qnum <=6
     store time() to starttime
     @1,0 clear
     @2,0 say "SET A - Information Request #"
     @2,29 say qnum
     text


     Please complete the DECISION WORKSHEET for the current
information
     request before beginning to enter your answers into the
computer.





     Press any key to continue when you are ready to enter your
endtext
     wait "    answers into the computer ..."
     clear
     store time() to midtime
     do SORT
     clear
     PUBLIC a,b,c,d,e,f,g,h
     store " " to a,b,c,d,e,f,g,h
     PUBLIC searchcond
     do SELECT
     clear
     if condition1 = " "
          searchcond = "    "
     endif
     if condition1 <> " " .and. condition2 =" "
          searchcond = "&condition1"
     endif
     if condition2 <> " "
          do LOGICAL
          store
a+condition1+b+LO1+c+condition2+d+LO2+e+condition3+f+LO3+g+con-
dition4+h to searchcond
     endif
```

```
      clear
      do ESSENT
      clear
      do VALIDATE
      set talk off
      clear
      store time() to endtime
      totaltime = (val(substr(endtime,1,2))*60*60 +
val(substr(endtime,4,2))*60;
      +val(substr(endtime,7,2))) - ((val(substr(starttime,1,2))*60*60;
      + val(substr(starttime,4,2))*60 + val(substr(starttime,7,2))))
      decidetime = (val(substr(midtime,1,2))*60*60 +
val(substr(midtime,4,2))*60;
      +val(substr(midtime,7,2))) - ((val(substr(starttime,1,2))*60*60;
      + val(substr(starttime,4,2))*60 + val(substr(starttime,7,2))))
      entertime = (val(substr(endtime,1,2))*60*60 +
val(substr(endtime,4,2))*60;
      +val(substr(endtime,7,2))) - ((val(substr(midtime,1,2))*60*60;
      + val(substr(midtime,4,2))*60 + val(substr(midtime,7,2))))
      do RETRIEVA
      clear
      set index to
      qnum = qnum + 1
enddo
set talk on
set status on
return
```

INTRO.PRG

```
clear
set talk off
set status off
text

                RETRIEVING INFORMATION FROM A DATABASE
                -------------------------------------

      During this set of activities you will be retrieving information
      from the database in response to 6 different information requests.


endtext
@20,0 say " "
wait
clear
text

      You will be asked to specify the following:
```

1. Does the data set need to be sorted. If "yes", which field should be the key field? Should the data be sorted in ascending or descending order?

2. Do selection criteria have to be specified? If "yes", which field names, operators, and arguments are needed to define each selection criterion? If there are two or more selection criteria, which logical operator(s) -- AND or OR -- are needed to link the selection criteria?

   Do parentheses need to be used to group selection criteria?

3. Which fields are "essential" fields -- fields that must be displayed to provide the required information?

4. Which fields are "validating" fields -- fields that allow the user to evaluate the effectiveness of the search strategy

```
endtext
wait
return
```

SORT.PRG

```
clear
set talk off
set bell off
set status off
set scoreboard off
@0,1 say "Information Request #"
@0,22 say qnum
@0,58 say "DEFINING SORT CRITERIA"
@2,0 clear
PUBLIC sort, ndxname, sortflag
store " " to sort
store 0 to sortflag
store "        " to ndxname
do while upper(sort) <> "N" .and. upper(sort) <> "Y"
     store " " to sort
     ?
     ?
     @10,5 say "     Does the data set, or the retrieved information,
need"
     do while upper(sort) <> "N" .and. upper(sort) <> "Y"
          @11,0 clear
          accept "          to be arranged in order (Y/N)? " to sort
     enddo
enddo
if upper(sort) = "Y"
     PUBLIC sortflag
     sortflag = 1
```

```
endif
@2,0 clear
if sortflag = 1
      @3,10 say "Which field should be used as the key field (the field
by"
      @4,10 say "which the data is arranged in order)?"
      @6,10 say "Select the key field by typing the number of the
field from"
      @7,10 say "the list below:"
      @9,10 say " 1   COUNTRY                    11   PER CAPITA
INCOME   "
     @10,10 say " 2   CONTINENT                  12   NUMBER OF
RADIO SETS"
     @11,10 say " 3   CAPITAL CITY               13   NUMBER OF
TV SETS"
     @12,10 say " 4   CURRENCY                   14   DAILY
NEWSPAPER CIRCULATION"
     @13,10 say " 5   LANGUAGES                  15   MALE LIFE
EXPECTANCY"
     @14,10 say " 6   TYPE OF GOVERNMENT         16   FEMALE
LIFE EXPECTANCY"
     @15,10 say " 7   MAJOR INDUSTRIES           17   LITERACY
RATE"
     @16,10 say " 8   GROSS DOMESTIC PRODUCT     18
CONSUMER PRICE INDEX"
     @17,10 say " 9   VALUE OF EXPORTS           19
POPULATION"
     @18,10 say "10   COST OF IMPORTS            20   RATE OF
POPULATION INCREASE"

     store "  " to key
     do while val(key) < 1 .or. val(key) > 20
          @21,0 clear
          accept "        Enter the number corresponding to the
key field: " to key
     enddo
     store val(key) to keyfield
     DO CASE
          CASE keyfield = 1
               kfname = "COUNTRY"
          CASE keyfield = 2
               kfname = "CONTINENT"
          CASE keyfield = 3
               kfname = "CAPITAL"
          CASE keyfield = 4
               kfname = "CURRENCY"
          CASE keyfield = 5
               kfname = "LANGUAGES"
          CASE keyfield = 6
               kfname = "GOVT_TYPE"
          CASE keyfield = 7
               kfname = "MAJ_INDUST"
          CASE keyfield = 8
               kfname = "GROSS_DP"
```

```
          CASE keyfield = 9
               kfname = "EXPORTS"
          CASE keyfield = 10
               kfname = "IMPORTS"
          CASE keyfield = 11
               kfname = "INCOME_PC"
          CASE keyfield = 12
               kfname = "RADIOS"
          CASE keyfield = 13
               kfname = "TVS"
          CASE keyfield = 14
               kfname = "DNC"
          CASE keyfield = 15
               kfname = "MLE"
          CASE keyfield = 16
               kfname = "FLE"
          CASE keyfield = 17
               kfname = "LIT_LEVEL"
          CASE keyfield = 18
               kfname = "CPI"
          CASE keyfield = 19
               kfname = "POPULATION"
          CASE keyfield = 20
               kfname = "POP_INC"
     ENDCASE
     @2,0 clear
     @10,10 say "Should the records be arranged in:"
     @12,15 say "1   Ascending order"
     @13,15 say "2   Descending order"
     PUBLIC order
     store " " to order
     do while order <> "1" .and. order <> "2"
          @15,0 clear
          accept "          Type 1 or 2 to select order of records: "
to order
     enddo
     if order = "1"
          if keyfield >=10 .and. keyfield <= 20
               ndxname = "UP"+str(keyfield,2)
          else
               ndxname = "UP"+str(keyfield,1)
          endif
     endif
     if order = "2"
          if keyfield >=10 .and. keyfield <= 20
               ndxname = "DOWN"+str(keyfield,2)
          else
               ndxname = "DOWN"+str(keyfield,1)
          endif
     endif
endif
return
```

SELECT.PRG

```
clear
set talk off
set bell off
set status off
set scoreboard off
@0,1 say "Information Request #"
@0,22 say qnum
@0,53 say "DEFINING SELECTION CRITERIA"
@2,0 clear
store " " to condition
do while upper(condition)<>"Y" .and. upper(condition)<>"N"
     @5,0 clear
     @6,10 say "Do selection criteria need to be defined so that the
correct"
     accept "          information can be retrieved for this
information request? " to condition
enddo
@2,0 clear
PUBLIC snum, snums
snum = 0
snums = 0
if upper(condition)="Y"
     @5,10 say "How many selection criteria do you wish to define?"
     do while snums < 1 .or. snums > 4
          @6,0 clear
          input "          Enter a number (1 - 4): " to snums
     enddo
     store 1 to snum
     store space(30) to sdname
     store space(10) to sname
     @2,0 clear
endif
PUBLIC condition1,condition2,condition3
PUBLIC condition4,condition5,condition6
store space(40) to condition1,condition2,condition3
store space(40) to condition4,condition5,condition6
do while upper(condition) = "Y".and.snum <= snums
      @2,0 clear
          @3,10 say "Which field name needs to be specified in
SELECTION "
          @4,10 say "CRITERION #"+ltrim(str(snum))+"?"
          @6,10 say "In the space provided below, type in the number
of the field"
          @7,10 say "from the following list:"
          @9,10 say " 1   COUNTRY              11   PER
CAPITA INCOME  "
          @10,10 say " 2   CONTINENT           12   NUMBER
OF RADIO SETS /1000"
          @11,10 say " 3   CAPITAL CITY        13   NUMBER
OF TV SETS /1000"
          @12,10 say " 4   CURRENCY            14   DAILY
NEWSPAPER CIRCULATION /1000"
```

```
        @13,10 say " 5  LANGUAGES              15  MALE
LIFE EXPECTANCY"
        @14,10 say " 6  TYPE OF GOVERNMENT      16
FEMALE LIFE EXPECTANCY"
        @15,10 say " 7  MAJOR INDUSTRIES       17
LITERACY RATE"
        @16,10 say " 8  GROSS DOMESTIC PRODUCT  18  %
CHANGE IN CONSUMER PRICE INDEX"
        @17,10 say " 9  VALUE OF EXPORTS       19
POPULATION"
        @18,10 say "10  COST OF IMPORTS        20  % RATE
OF POPULATION INCREASE"

        store 0 to sfield
        do while sfield < 1 .or. sfield > 20
            @20,0 clear
            input "                 Enter the number for the
field: " to sfield
        enddo
    DO CASE
        CASE sfield = 1
            sname = "COUNTRY   "
            sdname = "COUNTRY                    "
        CASE sfield = 2
            sname = "CONTINENT "
            sdname = "CONTINENT                  "
        CASE sfield = 3
            sname = "CAPITAL   "
            sdname = "CAPITAL CITY               "
        CASE sfield = 4
            sname = "CURRENCY  "
            sdname = "CURRENCY                   "
        CASE sfield = 5
            sname = "LANGUAGES "
            sdname = "LANGUAGES                  "
        CASE sfield = 6
            sname = "GOVT_TYPE "
            sdname = "TYPE OF GOVERNMENT         "
        CASE sfield = 7
            sname = "MAJ_INDUST"
            sdname = "MAJOR INDUSTRIES           "
        CASE sfield = 8
            sname = "GROSS_DP  "
            sdname = "GROSS DOMESTIC PRODUCT     "
        CASE sfield = 9
            sname = "EXPORTS   "
            sdname = "VALUE OF EXPORTS           "
        CASE sfield = 10
            sname = "IMPORTS   "
            sdname = "COST OF IMPORTS            "
        CASE sfield = 11
            sname = "INCOME_PC "
            sdname = "PER CAPITA INCOME          "
        CASE sfield = 12
```

```
                sname = "RADIOS    "
                sdname = "NUMBER OF RADIO SETS /1000       "
           CASE sfield = 13
                sname = "TVS       "
                sdname = "NUMBER OF TV SETS /1000          "
           CASE sfield = 14
                sname = "DNC       "
                sdname = "DAILY NEWSPAPER CIRCULATION /1000"
           CASE sfield = 15
                sname = "MLE  .    "
                sdname = "MALE LIFE EXPECTANCY             "
           CASE sfield = 16
                sname = "FLE       "
                sdname = "FEMALE LIFE EXPECTANCY           "
           CASE sfield = 17
                sname = "LIT_LEVEL "
                sdname = "LITERACY RATE                    "
           CASE sfield = 18
                sname = "CPI       "
                sdname = "% CHANGE IN CONSUMER PRICE INDEX "
           CASE sfield = 19
                sname = "POPULATION"
                sdname = "POPULATION                       "
           CASE sfield = 20
                sname = "POP_INC   "
                sdname = "% RATE OF POPULATION INCREASE    "
      ENDCASE
      @2,0 clear
      operator = 0
      @5,10 say "Select the relational operator for this selection
criterion"
      @6,10 say "from the list below:"
       @8,30 say "1   =   Equal to"
       @9,30 say "2   <=  Less than or equal to"
      @10,30 say "3   <   Less than"
      @11,30 say "4   >   Greater than"
      @12,30 say "5   >=  Greater than or equal to"
      @13,30 say "6   <>  Not equal to"
      @14,30 say "7   contains"
      do while operator < 1 .or. operator > 7
           @16,0 clear
           input "          Type the number (1 - 7): " to operator
      enddo
      store space(8) to op
      do case
           case operator = 1
                op = " = "
           case operator = 2
                op = " <= "
           case operator = 3
                op = " < "
           case operator = 4
                op = " > "
           case operator = 5
```

```
            op = " >= "
        case operator = 6
            op = " <> "
        case operator = 7
            op = " contains "
    endcase
    @2,0 clear
        @5,10 say "Enter the value or text in the space provided
below"
        @6,10 say "for the argument that will complete this
selection"
        @7,10 say "criterion."
        @10,5 say "&sdname"+" "+"&op"
        store space(22) to argument
    if sfield >= 8 .and. sfield <= 20
        @15,5 say "When entering an argument for a field that
contains only numbers,"
        @16,5 say "remember that ONLY DIGITS, and a DECIMAL
POINT (if needed) should"
        @17,5 say " be entered."
        @19,5 say "Do not include dollar signs, commas, percentage
signs, etc. when"
        @20,5 say "entering a numeric argument."
    endif
    if sfield >= 1 .and. sfield <= 7
        @10,50 get argument
        read
    else
        if argument = " "
            @10,50 get argument
            read
        endif
        do while
at(',',trim(argument))<>0.or.at('$',trim(argument))<>0.or.;
            at('%',trim(argument))<>0.or.argument=" "
            @10,50
            store space(22) to argument
            @10,50 get argument
            read
        enddo
    endif
        if sfield <= 7
            conarg = "'"+upper(trim(argument))+"'"
        else
            conarg = val(trim(argument))
        endif
    do case
        case snum = 1
            if operator = 7
                store conarg+" "+"$("+"upper("+trim(sname)+")"+")"
to condition1
            else
                if sfield <= 7
                    store
```

```
"upper("+trim(sname)+")"+trim(op)+conarg to condition1
                    else
                          store trim(sname)+trim(op)+str(conarg) to
condition1
                    endif
               endif
          case snum = 2
               if operator = 7
                    store conarg+" "+"$("+"upper("+trim(sname)+")"+")"
to condition2
               else
                    if sfield <= 7
                          store
"upper("+trim(sname)+")"+trim(op)+conarg to condition2
                    else
                          store trim(sname)+trim(op)+str(conarg) to
condition2
                    endif
               endif
          case snum = 3
               if operator = 7
                    store conarg+" "+"$("+"upper("+trim(sname)+")"+")"
to condition3
               else
                    if sfield <= 7
                          store
"upper("+trim(sname)+")"+trim(op)+conarg to condition3
                    else
                          store trim(sname)+trim(op)+str(conarg) to
condition3
                    endif
               endif
          case snum = 4
               if operator = 7
                    store conarg+" "+"$("+"upper("+trim(sname)+")"+")"
to condition4
               else
                    if sfield <= 7
                          store
"upper("+trim(sname)+")"+trim(op)+conarg to condition4
                    else
                          store trim(sname)+trim(op)+str(conarg) to
condition4
                    endif
               endif
          case snum = 5
               if operator = 7
                    store conarg+" "+"$("+"upper("+trim(sname)+")"+")"
to condition5
               else
                    if sfield <= 7
                          store
"upper("+trim(sname)+")"+trim(op)+conarg to condition5
                    else
```

```
                              store trim(sname)+trim(op)+str(conarg) to
condition5
                         endif
                    endif
               case snum = 6
                    if operator = 7
                         store conarg+" "+"$("+"upper("+trim(sname)+")"+")"
to condition6
                    else
                         if sfield <= 7
                              store
"upper("+trim(sname)+")"+trim(op)+conarg to condition6
                         else
                              store trim(sname)+trim(op)+str(conarg) to
condition6
                         endif
                    endif
          endcase
                    PUBLIC cond1
                    PUBLIC cond2
                    PUBLIC cond3
                    PUBLIC cond4
                    PUBLIC cond5
                    PUBLIC cond6
          do case
               case snum = 1
                    store trim(sdname)+trim(op)+" "+trim(argument) to cond1

               case snum = 2
                    store trim(sdname)+trim(op)+" "+trim(argument) to cond2

               case snum = 3
                    store trim(sdname)+trim(op)+" "+trim(argument) to cond3

               case snum = 4
                    store trim(sdname)+trim(op)+" "+trim(argument) to cond4

               case snum = 5
                    store trim(sdname)+trim(op)+" "+trim(argument) to cond5

               case snum = 6
                    store trim(sdname)+trim(op)+" "+trim(argument) to cond6

          endcase
          snum = snum + 1
     enddo
     do case
          case snums = 1
               condition2 =" "
               cond2 = " "
               condition3 =" "
               cond3 = " "
               condition4 =" "
               cond4 = " "
```

```
            condition5 =" "
            cond5 = " "
            condition6 =" "
            cond6 = " "
      case snums = 2
            condition3 =" "
            cond3 = " "
            condition4 =" "
            cond4 = " "
            condition5 =" "
            cond5 = " "
            condition6 =" "
            cond6 = " "
      case snums = 3
            condition4 =" "
            cond4 = " "
            condition5 =" "
            cond5 = " "
            condition6 =" "
            cond6 = " "
endcase
set talk on
return
```

LOGICAL.PRG

```
clear
set talk off
set bell off
set status off
set scoreboard off
@0,1 say "Information Request #"
@0,22 say qnum
@0,53 say "DEFINING LOGICAL OPERATORS "
@2,0 clear

set talk off
text



      When more than one selection criterion has been used to select
records
      from a database, the logical operators -- AND and OR -- must be
used
      to combine the selection criteria.



      Your next task is to select the logical operators to combine the
      selection criteria that you have defined. After you have selected
      the logical operators, you will be given the opportunity to
further
```

group some selection criteria within parentheses.

```
endtext
PUBLIC LO1,LO2,LO3
store "     " to LO1, LO2, LO3
store 0 to L1,L2,L3
wait
@2,0 clear
if condition2 > " "
      @2,10 say "&cond1"
      @6,10 say "&cond2"
@4,41 say "1: AND   2: OR "
if condition3 > " "
      @10,10 say "&cond3"
      @8,41 say "1: AND   2: OR "
endif
if condition4 > " "
      @14,10 say "&cond4"
      @12,41 say "1: AND   2: OR "
endif
if condition1 > " "
      do while L1< 1 .or. L1 > 2
            @4,15 get L1 picture '99' range 1,2
            read
      enddo
endif
if condition3 > " "
      do while L2< 1 .or. L2 > 2
            @8,15 get L2 picture '99' range 1,2
            read
      enddo
endif
if condition4 > " "
      do while L3< 1 .or. L3 > 2
            @12,15 get L3 picture '99' range 1,2
            read
      enddo
endif
if condition5 > " "
      do while L4< 1 .or. L4 > 2
            @16,15 get L4 picture '99' range 1,2
            read
      enddo
endif
if condition6 > " "
      do while L5< 1 .or. L5 > 2
            @20,15 get L5 picture '99' range 1,2
            read
      enddo
endif

if L1 = 1
```

```
     LO1 = " .AND. "
else
     if condition2 > " "
          LO1 = " .OR. "
     else
          LO1 = ""
     endif
endif
if L2 = 1
     LO2 = " .AND. "
else
     if condition3 > " "
          LO2 = " .OR. "
     else
          LO2 = ""
     endif
endif
if L3 = 1
     LO3 = " .AND. "
else
     if condition4 > " "
          LO3 = " .OR. "
     else
          LO3 = ""
     endif
endif
@2,0 clear
store " " to choice
do while condition3 <> " " .and.lower(choice)<> "z"
     text


     Following these instructions, the selection criteria that you have
     defined will be displayed, together with the logical operators AND

     and OR.

     In some search statements, parentheses have to be used to group
     criteria so that the information that we require is retrieved
     from the database.


     endtext
     wait
     clear
     @0,1 say "Information Request #"
     @0,22 say qnum
     @0,48 say "DEFINING POSITION OF PARENTHESES"
     @2,0 clear

     text
```

Look carefully at your selection criteria, then decide where -- if at all -- parentheses have to be included. Every position where a parenthesis can be placed is marked by a letter, A - H.

To include a parenthesis, type the letter that marks the position where the parenthesis is to be included. You will be asked if you want that position to show:

1 an opening parenthesis
2 a closing parenthesis
3 no parenthesis (this allows you to alter a previous decision)

Make sure that, if you use any parentheses, you have an equal number of opening and closing parentheses.

***** Many search statements don't require parentheses, so don't *****

think that you must include some.

```
endtext
wait
@2,0 clear
@2,5 say "*a*"
@2,10 say "&cond1"
@2,(12+len(cond1)) say "*b*"
@4,5 say "*c*"
@4,10 say "&cond2"
@4,(12+len(cond2)) say "*d*"
@3,15 say "&LO1"
if condition3 > " "
      @6,5 say "*e*"
      @6,10 say "&cond3"
      @6,(12+len(cond3)) say "*f*"
      @5,15 say "&LO2"
endif
if condition4 > " "
      @8,5 say "*g*"
      @8,10 say "&cond4"
      @8,(12+len(cond4)) say "*h*"
      @7,15 say "&LO3"
endif
store " " to choice
do while lower(choice) <> "z"
      @15,5 say "Enter the letter a - h to include or remove
parentheses."
      do while
(lower(choice)<"a".or.lower(choice)>"h").and.lower(choice) <> "z"
            @16,0 clear
            accept "    Enter z when you have finished.   " to
choice
      enddo
```

```
if lower(choice) <> "z"
     @18,5 say "Do you wish to:"
     @20,10 say "1  include an OPENING parenthesis"
     @21,10 say "2  include a CLOSING parenthesis"
     @22,10 say "3  remove a parenthesis from that
position"
     store 0 to pchoice
     do while pchoice < 1 .or. pchoice > 3
          @22,60 clear
          input "
 " to pchoice
     enddo
     do case
          case lower(choice) = "a"
               do case
                    case pchoice = 1
                         @2,5 say " ( "
                         a = "("
                    case pchoice = 2
                         @2,5 say " ) "
                         a = ")"
                    case pchoice = 3
                         @2,5 say "*a*"
                         a = ""
               endcase
          case lower(choice) = "b"
               do case
                    case pchoice = 1
                         @2,(12+len(cond1)) say " ( "
                         b = "("
                    case pchoice = 2
                         @2,(12+len(cond1)) say " ) "
                         b = ")"
                    case pchoice = 3
                         @2,(12+len(cond1)) say "*b*"
                         b = ""
               endcase
          case lower(choice) = "c"
               do case
                    case pchoice = 1
                         @4,5 say " ( "
                         c = "("
                    case pchoice = 2
                         @4,5 say " ) "
                         c = ")"
                    case pchoice = 3
                         @4,5 say "*c*"
                         c = ""
               endcase
          case lower(choice) = "d"
               do case
                    case pchoice = 1
                         @4,(12+len(cond2)) say " ( "
                         d = "("
```

```
            case pchoice = 2
                @4,(12+len(cond2)) say " ) "
                d = ")"
            case pchoice = 3
                @4,(12+len(cond2)) say "*d*"
                d = ""
        endcase
    case lower(choice) = "e"
        do case
            case pchoice = 1
                @6,5 say " ( "
                e = "("
            case pchoice = 2
                @6,5 say " ) "
                e = ")"
            case pchoice = 3
                @6,5 say "*e*"
                e = ""
        endcase
    case lower(choice) = "f"
        do case
            case pchoice = 1
                @6,(12+len(cond3)) say " ( "
                f = "("
            case pchoice = 2
                @6,(12+len(cond3)) say " ) "
                f = ")"
            case pchoice = 3
                @6,(12+len(cond3)) say "*f*"
                f = ""
        endcase
    case lower(choice) = "g"
        do case
            case pchoice = 1
                @8,5 say " ( "
                g = "("
            case pchoice = 2
                @8,5 say " ) "
                g = ")"
            case pchoice = 3
                @8,5 say "*g*"
                g = ""
        endcase
    case lower(choice) = "h"
        do case
            case pchoice = 1
                @8,(12+len(cond4)) say " ( "
                h = "("
            case pchoice = 2
                @8,(12+len(cond4)) say " ) "
                h = ")"
            case pchoice = 3
                @8,(12+len(cond4)) say "*h*"
                h = ""
```

```
                              endcase
                      endcase
              store " " to choice
              endif
         enddo
    enddo
    return
```

ESSENT.PRG

```
clear
set talk off
set bell off
set status off
set scoreboard off
@0,1 say "Information Request #"
@0,22 say qnum
@0,53 say "DEFINING ESSENTIAL FIELDS  "
@2,0 clear
store 1 to enum
store space(10) to ename1,ename2,ename3,ename4,ename5,ename6
store " " to essent
store "Y" to essent
do while upper(essent) = "Y"
     @2,0 clear
     @3,10 say "Which field should be displayed as one of the
ESSENTIAL "
     @4,10 say "FIELDS when the information is retrieved?"
     @6,10 say "In the space provided below, type in the number
corresonding"
     @7,10 say "to the eassential field from the following list:"
     @9,10 say " 1   COUNTRY                    11   PER CAPITA
INCOME  "
     @10,10 say " 2   CONTINENT                 12   NUMBER OF
RADIO SETS"
     @11,10 say " 3   CAPITAL CITY              13   NUMBER OF
TV SETS"
     @12,10 say " 4   CURRENCY                  14   DAILY
NEWSPAPER CIRCULATION"
     @13,10 say " 5   LANGUAGES                 15   MALE LIFE
EXPECTANCY"
     @14,10 say " 6   TYPE OF GOVERNMENT        16   FEMALE
LIFE EXPECTANCY"
     @15,10 say " 7   MAJOR INDUSTRIES          17   LITERACY
RATE"
     @16,10 say " 8   GROSS DOMESTIC PRODUCT   18
CONSUMER PRICE INDEX"
     @17,10 say " 9   VALUE OF EXPORTS          19
POPULATION"
     @18,10 say "10   COST OF IMPORTS           20   RATE OF
POPULATION INCREASE"

     store 0 to efield
```

```
do while efield < 1 .or. efield > 20
     @20,0 clear
     input "Enter the number corresponding to the essential field:
" to efield
enddo
store "ename"+str(enum,1) to ename
DO CASE
     CASE efield = 1
          &ename = "COUNTRY"
     CASE efield = 2
          &ename = "CONTINENT"
     CASE efield = 3
          &ename = "CAPITAL"
     CASE efield = 4
          &ename = "CURRENCY"
     CASE efield = 5
          &ename = "LANGUAGES"
     CASE efield = 6
          &ename = "GOVT_TYPE"
     CASE efield = 7
          &ename = "MAJ_INDUST"
     CASE efield = 8
          &ename = "GROSS_DP"
     CASE efield = 9
          &ename = "EXPORTS"
     CASE efield = 10
          &ename = "IMPORTS"
     CASE efield = 11
          &ename = "INCOME_PC"
     CASE efield = 12
          &ename = "RADIOS"
     CASE efield = 13
          &ename = "TVS"
     CASE efield = 14
          &ename = "DNC"
     CASE efield = 15
          &ename = "MLE"
     CASE efield = 16
          &ename = "FLE"
     CASE efield = 17
          &ename = "LIT_LEVEL"
     CASE efield = 18
          &ename = "CPI"
     CASE efield = 19
          &ename = "POPULATION"
     CASE efield = 20
          &ename = "POP_INC"
ENDCASE
@2,0 clear
store " " to essent
do while upper(essent) <> "Y" .and. upper(essent) <> "N"
     @10,0 clear
     store " " to essent
     accept "Do you need to identify another essential field? " to
```

```
essent
      enddo
      if upper(essent) = "Y"
            enum = enum + 1
      endif
enddo
PUBLIC efields
store
trim(trim(ename1)+","+trim(ename2)+","+trim(ename3)+","+trim(ename4);
+","+trim(ename5)+","+trim(ename6)) to efields
if right("&efields",5) = ",,,,,"
      temp = left("&efields",(len("&efields")-5))
      efields = "&temp"
endif
if right("&efields",4) = ",,,,"
      temp = left("&efields",(len("&efields")-4))
      efields = "&temp"
endif
if right("&efields",3) = ",,,"
      temp = left("&efields",(len("&efields")-3))
      efields = "&temp"
endif
if right("&efields",2) = ",,"
      temp = left("&efields",(len("&efields")-2))
      efields = "&temp"
endif
if right("&efields",1) = ","
      temp = left("&efields",(len("&efields")-1))
      efields = "&temp"
endif
set talk on
return
```

VALIDATE.PRG

```
clear
set talk off
set bell off
set status off
set scoreboard off
@0,1 say "Information Request #"
@0,22 say qnum
@0,53 say "DEFINING VALIDATING FIELDS "
@2,0 clear
store 1 to vnum
store space(10) to vname1,vname2,vname3,vname4,vname5,vname6
store " " to valid
store "Y" to valid
do while upper(valid) = "Y"
      @2,0 clear
      @3,10 say "Which field should be displayed as one of the
VALIDATING "
      @4,10 say "FIELDS when the information is retrieved?"
```

```
@6,10 say "In the space provided below, type in the number
corresponding"
@7,10 say "to the validating field from the following list:"
@9,10 say " 1  COUNTRY                 11  PER CAPITA
INCOME  "
@10,10 say " 2  CONTINENT              12  NUMBER OF
RADIO SETS"
@11,10 say " 3  CAPITAL CITY           13  NUMBER OF
TV SETS"
@12,10 say " 4  CURRENCY               14  DAILY
NEWSPAPER CIRCULATION"
@13,10 say " 5  LANGUAGES              15  MALE LIFE
EXPECTANCY"
@14,10 say " 6  TYPE OF GOVERNMENT       16  FEMALE
LIFE EXPECTANCY"
@15,10 say " 7  MAJOR INDUSTRIES       17  LITERACY
RATE"
@16,10 say " 8  GROSS DOMESTIC PRODUCT  18
CONSUMER PRICE INDEX"
@17,10 say " 9  VALUE OF EXPORTS       19
POPULATION"
@18,10 say "10  COST OF IMPORTS        20  RATE OF
POPULATION INCREASE"

store 0 to vfield
do while vfield < 1 .or. vfield > 20
     @20,0 clear
     input "Enter the number corresponding to the validating
field: " to vfield
enddo
store "vname"+str(vnum,1) to vname
DO CASE
     CASE vfield = 1
          &vname = "COUNTRY"
     CASE vfield = 2
          &vname = "CONTINENT"
     CASE vfield = 3
          &vname = "CAPITAL"
     CASE vfield = 4
          &vname = "CURRENCY"
     CASE vfield = 5
          &vname = "LANGUAGES"
     CASE vfield = 6
          &vname = "GOVT_TYPE"
     CASE vfield = 7
          &vname = "MAJ_INDUST"
     CASE vfield = 8
          &vname = "GROSS_DP"
     CASE vfield = 9
          &vname = "EXPORTS"
     CASE vfield = 10
          &vname = "IMPORTS"
     CASE vfield = 11
          &vname = "INCOME_PC"
```

```
            CASE vfield = 12
                &vname = "RADIOS"
            CASE vfield = 13
                &vname = "TVS"
            CASE vfield = 14
                &vname = "DNC"
            CASE vfield = 15
                &vname = "MLE"
            CASE vfield = 16
                &vname = "FLE"
            CASE vfield = 17
                &vname = "LIT_LEVEL"
            CASE vfield = 18
                &vname = "CPI"
            CASE vfield = 19
                &vname = "POPULATION"
            CASE vfield = 20
                &vname = "POP_INC"
        ENDCASE
        @2,0 clear
        store " " to valid
        @10,10 say "Do you need to identify another validating field? "
        store " " to valid
        do while upper(valid) <> "Y" .and. upper(valid) <> "N"
            @10,0 clear
            store " " to valid
            accept "Do you need to identify another validating field? "
to valid
        enddo
        if upper(valid) = "Y"
            vnum = vnum + 1
        endif
enddo
PUBLIC vfields
store
","+trim(vname1)+","+trim(vname2)+","+trim(vname3)+","+trim(vname4);
+","+trim(vname5)+","+trim(vname6) to vfields
if right("&vfields",5) = ",,,,,"
        temp = left("&vfields",(len("&vfields")-5))
        vfields = "&temp"
endif
if right("&vfields",4) = ",,,,"
        temp = left("&vfields",(len("&vfields")-4))
        vfields = "&temp"
endif
if right("&vfields",3) = ",,,"
        temp = left("&vfields",(len("&vfields")-3))
        vfields = "&temp"
endif
if right("&vfields",2) = ",,"
        temp = left("&vfields",(len("&vfields")-2))
        vfields = "&temp"
endif
if right("&vfields",1) = ","
```

```
      temp = left("&vfields",(len("&vfields")-1))
      vfields = "&temp"
 endif
 return
```

RETRIEVA.PRG

```
clear
set talk off
set bell off
set status off
set scoreboard off
@0,1 say "Information Request #"
@0,22 say qnum
@0,58 say "RETRIEVING INFORMATION"
@2,0 clear
set talk off
PUBLIC search
if ltrim(searchcond) > " "
      store "LIST OFF "+"&EFIELDS"+"&VFIELDS"+" FOR
"+"&SEARCHCOND" to search
else
      store "LIST OFF NEXT 6 "+"&EFIELDS"+"&VFIELDS" to search
endif
text


    Your search statement is being executed. When you press any key
    to continue, you will see the retrieved information displayed
    on the screen.


endtext
use data
wait
@2,0 clear
set index to &ndxname
&search
?
WAIT "Press the SPACE BAR to see the answer for this request..."
DO CASE
      CASE qnum = 1
            set index to down8
            list next 6 COUNTRY, GROSS_DP OFF
            SET INDEX TO
      CASE qnum = 2
            LIST RADIOS,COUNTRY FOR COUNTRY="Afghanistan" OFF
      CASE qnum = 3
            SET INDEX TO UP1
            LIST COUNTRY, CPI,INCOME_PC FOR
CPI<5.AND.INCOME_PC>9000 OFF
            SET INDEX TO
```

```
      CASE qnum = 4
           LIST COUNTRY,CONTINENT,RADIOS,LIT_LEVEL FOR;
           CONTINENT<>"Asia".AND.RADIOS<300.AND.LIT_LEVEL>95
 OFF
      CASE qnum = 5
           SET INDEX TO UP1
           LIST COUNTRY,CONTINENT,CURRENCY FOR
CONTINENT="Europe".AND.;
           (CURRENCY="Pound".OR.CURRENCY="Mark") OFF
           SET INDEX TO
      CASE qnum = 6
           LIST COUNTRY,CONTINENT,POP_INC,LANGUAGES FOR
CONTINENT<>;
           "Europe".AND.POP_INC<=1.5.AND.("English"
$(LANGUAGES).OR.;
           "French" $(languages)) OFF

ENDCASE
store " " to continue
do while upper(continue)<>"C"
     @22,0 clear
     accept "Press the letter C, then press <Return> to continue ..." to
continue
enddo
store "SETA"+"_"+ltrim(str(qnum)) to searchfile
set alte to &searchfile
set alte on
set console off
? DATE,STARTTIME, MIDTIME, ENDTIME, TOTALTIME, DECIDETIME,
ENTERTIME
? SORT, NDXNAME
? SEARCHCOND
? EFIELDS
? VFIELDS
?
? SEARCH
IF SORTFLAG = 1
     SET INDEX TO &NDXNAME
ENDIF
&SEARCH
SET INDEX TO
?

set console on
set alte off
return
```