AN EMPIRICAL INVESTIGATION OF THE EFFECTS OF INDIVIDUAL DIFFERENCES AND DATA MODELS ON THE EASE-OF-USE OF DATABASE QUERY FACILITIES BY CASUAL USERS

DISSERTATION

Presented to the Graduate Council of the North Texas State University in Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

by

Howard N. Ray, B.S., M.B.A.

Denton, Texas

August, 1984
The problem motivating this study is that database query facilities are not effectively meeting the needs of casual users. A solution to this problem is especially important due to the increasing number of potential casual users. There is considerable controversy revolving around the question of which elements and/or which combination of elements within the casual users' environment are necessary to provide an effective man/machine interface. The purpose of the study is to extend the basic knowledge relating to the effect of using different data models, the effect of cognitive style differences, and the interaction effects of these two independent variables on casual users' performance and confidence in writing database queries. The data models being investigated are the relational, hierarchical, and network models. The cognitive style classifications used are two of the Jungian dimensions, Sensing/Intuition and Thinking/Feeling.

The data collection process was divided into two phases. Phase I was the administration of the MBTI to all subjects. The subjects from each personality group were then randomly selected and alternately assigned to one of the three data model groups. Phase II of the data collection process was the administration of an information retrieval exercise. The exercise was a written
test designed to permit the evaluation of the subject's ability to comprehend and retrieve information from a database by writing English-like queries. The subject's performance in writing the queries was based on errors made in (1) writing the specification portion, (2) writing the condition portion, (3) writing the navigation portion, and (4) use of the language. The subject's confidence was a self-reported value on a scale of one to five.

The group using the relational model performed significantly better than the group using the hierarchical model when writing the specification portion and when writing the navigation portion of the query. The Intuition types performed significantly better than the Sensing types in terms of "language use". The Thinking types were significantly more confident than the Feeling types. This was also supported by the fact that the NT types were more confident than the SF types. No significant differences in confidence level was detected among the data model groups. No interaction effects between data models and cognitive style classifications were detected.

The results of this study indicate that using the relational model to present the external view will increase casual user performance in writing database queries. Other studies have determined that performance and confidence is better when using the hierarchical model. However, these studies used non-query writing tasks. This could be an indication that some of the dependent variables being used to measure ease-of-use do not reflect the users confidence and/or ability to write database
queries. The study also indicates that cognitive style need not be considered in the selection of the data model used to present the external view. However, it is indicated that Feeling personality types need additional support during training to increase their confidence level.
TABLE OF CONTENTS

LIST OF TABLES ........................................ iii
LIST OF ILLUSTRATIONS ............................... iv
LIST OF APPENDICES ................................... v

Chapter

I. INTRODUCTION ....................................... 1

1.1 Introduction ........................................ 1
1.2 Statement of the Problem ........................... 3
1.3 Purpose of the Study ............................... 6
1.4 Background .......................................... 7
1.5 Significance of the Study ......................... 13

II. LITERATURE REVIEW ............................... 20

2.1 Related Research ................................. 20
   2.1.1 Data Model Studies .......................... 20
   2.1.2 Cognitive Style Studies ..................... 28
2.2 Needed Research ................................... 32
2.3 Definition of Terms ............................... 35

III. RESEARCH METHODOLOGY ....................... 40

3.1 Research Framework .............................. 40
3.2 Research Hypotheses ............................. 42
3.3 Research Design .................................. 47
3.4 Procedures for Collection of Data ............... 51
   3.4.1 Exercises Used ............................... 53
   3.4.2 Grading the Exercises ....................... 56
3.5 Procedures for Analyzing the Data ............... 61
3.6 Limitations ....................................... 67

IV. RESULTS AND ANALYSIS ......................... 72

4.1 Data Analysis ..................................... 72
4.2 Subject Demographics ............................. 73
4.3 Jungian Classifications .......................... 74
4.4 Descriptive Statistics ........................... 75
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>Research Variable Set</td>
<td>40</td>
</tr>
<tr>
<td>II.</td>
<td>Summary of ANOVA Tests</td>
<td>66</td>
</tr>
<tr>
<td>III.</td>
<td>Cell Frequencies for the Sensing/Intuition Dimension by Data Models</td>
<td>74</td>
</tr>
<tr>
<td>IV.</td>
<td>Cell Frequencies for the Thinking/Feeling Dimension by Data Models</td>
<td>75</td>
</tr>
<tr>
<td>V.</td>
<td>Cell Frequencies for Jungian Personality Types by Data Models</td>
<td>75</td>
</tr>
<tr>
<td>VI.</td>
<td>Mean Performance Score (Range 0 - 6) and Percentage of Subjects Making No errors</td>
<td>81</td>
</tr>
<tr>
<td>VII.</td>
<td>Frequency Distribution of Specification Performance Scores</td>
<td>78</td>
</tr>
<tr>
<td>VIII.</td>
<td>Frequency Distribution of Condition Performance Scores</td>
<td>78</td>
</tr>
<tr>
<td>IX.</td>
<td>Frequency Distribution of Navigation Performance Scores</td>
<td>79</td>
</tr>
<tr>
<td>X.</td>
<td>Frequency Distribution of Language Use Performance Scores</td>
<td>79</td>
</tr>
<tr>
<td>XI.</td>
<td>Frequency Distribution of Overall Performance Scores</td>
<td>80</td>
</tr>
<tr>
<td>XII.</td>
<td>Frequency Distribution of Subject's Confidence Level</td>
<td>81</td>
</tr>
<tr>
<td>XIII.</td>
<td>Summary of Significant Findings and Tests</td>
<td>94</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>1</td>
<td>Descriptive Model of the Casual User's Environment</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>Model of Jungian Personality Types</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>Sample Grade Sheet</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>Factorial Design Model 1</td>
<td>63</td>
</tr>
<tr>
<td>5</td>
<td>Factorial Design Model 2</td>
<td>64</td>
</tr>
<tr>
<td>6</td>
<td>Factorial Design Model 3</td>
<td>65</td>
</tr>
<tr>
<td>7</td>
<td>Specification Errors</td>
<td>83</td>
</tr>
<tr>
<td>8</td>
<td>Condition Errors</td>
<td>83</td>
</tr>
<tr>
<td>9</td>
<td>Navigation Errors</td>
<td>84</td>
</tr>
<tr>
<td>10</td>
<td>Language Use Errors</td>
<td>84</td>
</tr>
<tr>
<td>11</td>
<td>Overall Performance</td>
<td>85</td>
</tr>
<tr>
<td>12</td>
<td>User Confidence</td>
<td>85</td>
</tr>
<tr>
<td>13</td>
<td>ANOVA Cell Means and Corresponding p Values</td>
<td>86</td>
</tr>
</tbody>
</table>
# LIST OF APPENDICES

<table>
<thead>
<tr>
<th>APPENDIX</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Exercise A (Relational) -- Instruction Booklet</td>
<td>102</td>
</tr>
<tr>
<td>B. Exercise B (Hierarchical)</td>
<td>119</td>
</tr>
<tr>
<td>C. Exercise C (Network)</td>
<td>136</td>
</tr>
<tr>
<td>D. Information Retrieval Problems</td>
<td>153</td>
</tr>
<tr>
<td>E. Subject Background Information Sheets</td>
<td>157</td>
</tr>
<tr>
<td>F. Myers-Briggs Type Indicator</td>
<td>159</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

1.1 Introduction

This study evaluates the effects of different data models and differences in cognitive style on casual users' performance and confidence in writing database queries. The data models used in the experiment are the relational, the hierarchical, and the network models. The cognitive style variables are two of the Jungian personality dimensions, Thinking/Feeling and Sensing/Intuition. Two dependent variables (user performance and confidence) were chosen as measures of database query facilities success.

Research and development activities during the past ten years have lead to significant improvements in the ease-of-use of database query languages (2, 3, 4, 5, 7, 14, 18, 19, 21, 22). Most of these research efforts were concerned with controversial issues related to query language design. The query languages being investigated have ranged from very formal and structured languages to "natural" language interfaces and other artificial intelligence (AI) based conversational systems (9, 11).

Welty (22, p. 13) identifies the major issues affecting query facilities ease-of-use to be:

1. Natural language vs. artificial language
2. Specification (nonprocedural) vs. procedural languages
3. Linear keyword vs. two dimension positional languages
4. Data model selection
5. Hardware factors.

However, another potentially important aspect of the man/machine system is the individual characteristics of the human element. Zmud (23) examines individual differences by providing a synthesis of the research literature from a variety of disciplines. The studies reported in the Zmud paper measured the effect of a number of cognitive, personality, demographic, and situational variables upon the use of management information systems and decision behavior. Zmud concludes that cognitive characteristics of decision makers do impact their informational needs.

This study investigates the relationships among data models, cognitive style, and casual user performance in writing database queries correctly. The specific areas to be examined are (1) the effect of different data models (relational, hierarchical, network) used in presenting the external view of a database, (2) the effect of individual differences in users' cognitive styles as measured by the Myers-Briggs Type Indicator (MBTI), and (3) the interaction effects of these two independent variables upon user confidence and performance in a query writing exercise.
1.2 Statement of the Problem

The motivation for the study is that researchers are reporting that database query facilities are not meeting the needs of casual users (1, 3, 7, 11, 21, 22). A better understanding of the relationships among critical variables within the casual users' environment will aid developers in meeting their needs. Ineffective user/system interfaces are especially critical due to the increasing number of users and potential users falling into the "casual" classification. Cuff (3, p. 163) says, "Although the number of queries which any individual wishes to pose may be small, the casual user population may be high enough to make provisions for their needs an important matter."

The diversity of individual characteristics exhibited by casual users and the diversity of environments in which they exist contribute to the complexity of meeting their needs. This is especially true when compared to meeting the needs of those who use computer systems more frequently than casual users.

Characteristics of the casual user's environment that contribute to this problem are infrequent use of the system, design characteristics of the system, and the wide variety of users' personality types involved. Cuff (3) points out that infrequent use of query facilities results in (1) poor retention of details, (2) propensity for error, (3) need for safety nets, (4) reluctance to accept initial training, and
(5) limited understanding of database details. Cuff also suggests that personality variations result in (1) inconsistent need for and use of documentation, (2) intolerance of formality, (3) inability to handle logic construction, and (4) dialogue preferences. Nickerson (17) identifies several environmental factors resulting in the casual user's electing not to use query systems. These include terminal accessibility and availability, start-stop hassle, poor response time, and work-session interrupts. Nickerson also emphasizes two of Cuff's concerns—poor documentation and poor conceptualization of the inner workings of the system.

This study investigates the effectiveness of different data models used in presenting the external view of the database. The external view is a means by which the database designer can provide the user with an accurate description of the database contents. Hoffer (10, p. 154) says, "Today, a critical issue is whether the data/information resource, represented in end-user programming facilities and described in database documentation, is consistent with the cognitive, personality, and demographic variables of end-users, who are increasingly the direct information system user/builder." This concern was also shared by Codd (4, p. 179), who noted that:

The user's view of the data in a formatted database has a fundamental impact on the way he conceives
and formulates queries and other types of transactions. The user's data model clearly should not have a multiplicity of structural alternatives for representing data. Such a multiplicity is incompatible with the casual user's unwillingness to consciously engage in a learning process and with his tendency to forget what he may have learned unconsciously, because of the irregularity of his interactions.

Another factor which may effect the casual user's performance and/or confidence is the individual's cognitive style. Zmud (23, p. 975) in his literature review of empirical studies concluded that individual differences do exert a major force in determining management information systems (MIS) success. He reports that "With organizations committing an increasing portion of their MIS efforts toward the utilization of the computer to actively support decision-making, the potential payoff from further research investigating the relationships between MIS success and personal characteristics of MIS users is high."

The specific aspects of the problem upon which this study concentrates are the effects of different data models used in presenting the external view of the database and differences in cognitive styles. The effects of these two independent variables are measured in terms of casual user performance and confidence in an information retrieval exercise. The exercise involves writing English-like non-procedural queries to retrieve information from a sales-related database. This approach was chosen because (1) the language is consistent with the trend toward natural
language query systems, (2) it has been reported (22) that non-procedural languages may be more effective for less experienced users, and (3) the use of only one language for all experimental groups will permit more accurate determination of the effects of the different data models. The combination of data models and personality types resulting in the best performance in this simulated query facilities environment are determined. Both user performance and confidence are used as measures of database query facilities success.

1.3 Purpose of the Study

The purpose of this study is to extend the basic knowledge relating to database query facilities ease-of-use for casual users. The main objective is to investigate the effect of using different data models to present the external view of a database on casual users' performance and confidence in writing database queries. The secondary objective is to investigate the effect of personality differences on casual user performance and confidence, and the interaction effects between data models and personality types. These objectives are to be accomplished by answering the following research questions:

(1) How is casual user performance in writing database queries affected by the data model employed (hierarchical, network, or relational)?
(2) How is casual user confidence in writing database queries affected by the data model employed (hierarchical, network, or relational)?

(3) How is casual user performance in writing database queries affected by differences in cognitive style?

(4) How is casual user confidence in writing database queries affected by differences in cognitive style?

(5) What are the interaction effects of the three data models and the cognitive style dimensions on casual user performance?

(6) What are the interaction effects of the three data models and the cognitive style dimensions on casual user confidence?

1.4 Background

The proliferation of database management systems, communications networks, and low-cost computing hardware, in conjunction with growing end-user awareness, is resulting in an ever-increasing backlog of systems design requests. A recent survey of thirty-three mainframe sites conducted by Techtran Industries, Inc., reports that the average delay for application development is increasing at a rate of about three months per year, and that the average user would have to wait 3-1/2 years for development work to begin on an
application requested today (8). Many organizations are dealing with backlogs of systems services requests through fourth generation programming languages, user involvement in the design process, and end-user developed queries.

This study is concerned with the growing population of casual users and environmental factors which may effect database query development. Effective use of systems by less experienced users requires a user/system interface that is easy both to understand and to use (3, 7, 10). In support of this concern, Harris (9, p. 303) says, "it is nearly impossible for end-users to make use of computer systems, in spite of our best attempts at user-oriented languages and database design." Ehrenreich (7, p. 709) reports that:

Computers are increasingly being employed by unsophisticated users entering and retrieving information from large databases. This presents a challenge to the designers of new systems, since these users are less willing and less able to adapt to even moderately complex procedures than are more experienced users. Designers of new systems will have to accommodate the new class of users, rather than expecting users to adapt to the systems (as has been the practice in the past).

Also supporting this concern, Mozeico (15, p. 100) says, "although desk-top and similar small computers are now available, not always are potential users familiar with their operation. In fact, the number of new inexperienced users appear to be growing. All of this suggests that a reassessment of the man-machine interface is in order."

Shneiderman (21, p. 417) sums up the situation by saying:
As questions of technical feasibility and performance of database systems are resolved, increased attention is being paid to human factors. There is widespread recognition that future systems will be commercially viable only if the user interface is in harmony with user skills and task requirements.

Reisner (18) reports on a number of human factors studies that have attempted to define "ease-of-use" of formal database query languages in terms that permit quantitative measurement. The Reisner paper also provides an excellent summary of the types of tests and the types of tasks within these tests which have been employed to measure ease-of-use.

This vague notion of database query facilities ease-of-use for casual users includes a variety of environmental, mechanical, technical, psychological, and behavioral aspects. Figure 1 presents a descriptive model of the casual user's environment, the major components found within the environment, the interaction of these components, and some of the more influential variables associated with each component. Many of the variables included in this model were identified and discussed by Nickerson (17) and Cuff (3).

The model illustrates that there are six primary components to be considered if effective systems are to be realized. These include (1) characteristics of the user, (2) characteristics of the user's environment, (3) design characteristics of the query facilities,
(4) physical characteristics of the terminal device, (5) characteristics of the software system that supports the database, and (6) the organizational context within which the user is functioning.

The model in Figure 1 also identifies specific influential variables associated with each major component, most of which could be further subdivided. The problem of providing query facilities to support the most effective mix of these numerous variables is formidable.

One of the more controversial issues relating to database systems ease-of-use is the external data model (user's conceptualization or image of the underlying data). According to Nickerson (17, p. 480), "The importance of the issue of users' models has been stressed by Newman and Sproull (16), but very little effort has been devoted to the development of guidelines for the invention and use of such models. This is, I believe, a ripe area for research."

Also, because of the wide diversity of information systems users within a typical business organization, researchers are attempting to identify homogeneous groups of users who are able to effectively use systems with a given set of attributes. Keen and Bronsema (13, p. 21) report that "the link between cognitive style and the implementation and use of information systems and models is a recurring theme in MIS/MS research."

Much of the prior query language research did not
investigate individual differences, or the specific problem of ineffective "casual user" interfaces. Consequently, the research findings from these studies have limited applicability to the casual user problem. This limitation is primarily the result of

(1) Design characteristics resulting in the effect of the data models being confounded by the effect of the data manipulation languages.

(2) The experience level of the subjects was not representative of the experience level of the casual user population. One common characteristic of the casual user is his lack of familiarity with computer systems. However, subjects used in most of the previous experiments were either students who had completed a number of programming courses, or experienced computer programmers.

(3) There were external validity problems due to the selection of dependent variables. For example, the user's ability to memorize the database contents has been used as a surrogate for database success. However, it has not been demonstrated that the ability to memorize the contents of a database will have a positive correlation with one's ability to write database queries correctly.
(4) Another external validity problem in some experiments was that the database, or at least the external view used in the experiment, was unrealistically oversimplified.

This study is an extension of prior data model research (primarily Lockovsky [14], Hoffer [10], and Brosey and Shniederman [1]). It concentrates on psychological differences among casual users as measured by the MBTI, and determines the external data model, among the relational, hierarchical, and network models, that best supports effective query writing. It avoids the limitations outlined above by (1) using one English-like query language for all data models, (2) using subjects with little or no prior computer experience, (3) providing an experiment requiring the subjects to write database queries using a realistically involved external view.

1.5 Significance of the Study

One of the main objectives of Management Information Systems research is to improve the effectiveness of computer based information systems within the business environment (6). Effective direct use of database query facilities by business decision-makers will support this objective by (1) reducing the delay in making information available, (2) eliminating the possibility of an intermediary misinterpreting informational needs, (3) reinforcing the
user's notion of what he requires so that it becomes clearer to him, (4) adding flexibility and encouraging spontaneous ideas of different data arrangements or additional information to support the particular decision while permitting these ideas to be readily pursued, and (5) relieving some of the burden on the programming staff to provide ad hoc programs and queries.

Keen (11) says, "Interactive systems facilitate the use of the computer's analytical power and data retrieval capabilities, for managers as part of their ongoing decision-making process. The system software can be designed to allow a man/machine dialogue which transforms the computer into a personalized problem-solving tool."

Zmud (23), in his review of the MIS research literature, points out that MIS success is dependent upon a variety of factors—MIS policies, task characteristics, personal characteristics, organizational characteristics, and environmental factors, to name a few. The effectiveness of database query facilities designed to support casual users, because of their personal and environmental characteristics, requires interactive facilities that are easy to understand and use (3).

This study builds upon prior data model research in order to increase the accumulation of knowledge about this aspect of the man/machine interface. Based on the increasing size of the casual user population and the
potential casual user population, this study will make significant contribution to knowledge because it will provide an analysis of

1. the effect of each of the three commonly used data models on casual users' performance and confidence,

2. the effect of individual differences in cognitive style on casual users' performance and confidence, and

3. the interaction effects of data models and cognitive style on casual users' performance and confidence in writing database queries.

The findings from this study will extend existing knowledge about the relationships between controllable variables within the man/machine interface. These results will assist in the selection of documentation and training techniques which will provide a more compatible interface for this type of user.

A deeper understanding in these areas will be an aid to both software developers and user organizations. Designers and developers of future database query facilities will be able to improve the effectiveness of their software product and supporting documentation. End-user organizations will have better criteria for query facility selection and be able to provide the necessary environment for effective use.
The specific areas within the user's environment which are expected to receive the most benefits are user training, data model selection, and system documentation.
CHAPTER BIBLIOGRAPHY


17


CHAPTER II
LITERATURE REVIEW

2.1 Related Research

The presentation of the literature relating to the effect of different cognitive styles and data models on the ease-of-use of query facilities by casual users is divided into two areas: (1) database query facility studies dealing with some combination of data models, cognitive styles, and casual users, and (2) other MIS studies (non-query facilities) dealing with cognitive styles.

2.1.1 Data Model Studies

Most prior experimental studies of query systems characteristics, other than data models, have evaluated a variety of ease-of-use aspects by comparing two or more formal query languages (19). Although these studies have contributed to design improvements in the languages, it is believed by many that most commercially available query facilities are still inadequate to meet casual users needs (7, 9).

A recent study by Hoffer (11) was designed to analyze the way humans organize data when they are not forced to use a specific model. This experiment was conducted in an (open) laboratory environment using MBA students as subjects. The students were enrolled in an introductory MIS
course, and were made-up of both MIS and non-MIS majors. Prior to the experiment the subjects were classified on a number of demographic variables and four cognitive style dimensions which were based on the Klob Learning Style Inventory (LSI). Hoffer (11, p. 156) reports that the LSI "measures cognitive style along similar dimensions to the Myers-Briggs Type Indicator."

The subjects were then randomly assigned to one of four situation groups to insure an approximately equal number with the same cognitive style and experience level in each group. The groups were then given different organizational situations and asked to draw on paper the database needed to support certain functions.

The reported findings from this study suggest a strong preference (54.2%) for the use of flow diagrams in describing data. Other techniques used included table-like (20%), hierarchical (10%), and network (9%). One explanation offered by Hoffer (11) for this strong preference is "that people have difficulty, especially within a task context, separating data from process; possibly process gives meaning to data." A more logical explanation might be that, because of programming courses, flow diagrams were the only information system documentation technique that the subjects were comfortable in using. Hoffer cites this as one area needing further research.

The results of Hoffer's study suggested that cognitive
style, as measured by the Klob Learning Style Inventory, need not be considered in selecting a data model, but he does point out that other cognitive variables should be explored. User confidence was used as a surrogate of database success. The study found that increased user confidence was related to (1) more specific tasks, (2) greater programming experience, and (3) increased situation familiarity.

Another study by Lochovsky (17) consisted of two experiments designed to analyze user performance in writing database queries. These experiments involved the use of an Educational Data Base System (EDBS) developed and used at the University of Toronto. This system supports the use of all three commonly used data models. The data manipulation language (DML) used in support of each data model was designed to be similar to different formal query languages. The relational DML was similar to ALPHA, the hierarchical DML was similar to DL/1, and the network DML was similar to the DBTG DML.

The first experiment used forty-five computer science majors. Two thirds of these subjects were senior level undergraduates and the rest were graduate students. Twenty-three of the students worked as application programmers, the rest were classified as "computer users." The independent variables used in this experiment were programming experience, data models/DML, and different applications.
Three different applications were selected; each could be modelled more "naturally" using one of the three data models. The dependent variables were correctness of code, time to code, and time to debug. It was reported from this experiment that the relational group made fewer errors than the hierarchical group. No other significant differences were reported.

In the second experiment, sixteen business majors and forty-two computer science majors were used. The fifty-eight subjects were classified as beginning programmers (less than six months experience) or advanced programmers (more than six months of actual programming experience). The independent variables were programming experience level, data model/DML, and experience with the query language. The evaluation of query language experience was accomplished by administering two sets of queries. The first set was administered just after the subjects had learned the language, and the second set was administered after the first set had been debugged.

The reported results of the second experiment were that (1) beginning programmers using the relational model/DML performed significantly better than those using either of the other two models, on both sets of queries, (2) advanced programmers using the relational model/DML performed significantly better than those using either of the other two models, but only on the first set of problems.
Lochovsky does caution, however, that the effects of the data model is confounded by the differences in the DML, and that the results should not be interpreted as a general comparison of DBMS. One basic difference is that ALPHA is a much less procedural language than either DL/1 or DBTG DML.

Brosey and Shneiderman (3) conducted two experiments designed to compare the use of hierarchical versus relational data models. These experiments employed comprehension, problem-solving, and memorization tasks.

The first experiment used thirty-eight computer science undergraduates. These subjects were divided into two groups, advanced (six or more programming courses) and beginners (two to five programming courses). The results of this study were as follows:

(1) The advanced group performed significantly better on the comprehension task than the less experienced group when using the relational model.

(2) The less experienced group's comprehension was significantly better using the hierarchical model compared to the use of the relational model.

(3) No significant findings resulted from the problem-solving situation.

(4) A significant improvement in memorization was observed for groups using the hierarchical model.
Brosey and Shniederman's second experiment involved comprehension and memorization tasks using three models: relational (three relations), hierarchical-1 (two segments), and hierarchical-2 (two hierarchies). In the comprehension task, the performance was significantly better using the hierarchical-2 than the relational model. In the memorization task, significantly better performance occurred using the hierarchical-1 model than either the relational or the hierarchical-2 data model. In considering question complexity, the researchers found significantly better performance with the hierarchical-2 model over hierarchical-1 or relational, but only for complex questions. General conclusions from this study are that the hierarchical structure may be easier to use in some situations, especially for beginners. However, the database used in this experiment did have a "natural tree structure."

Another study, by Broadbent and Broadbent (2), involved the assignment and use of descriptor terms to identify 100 objects that might conceivably be given as Christmas presents. The purpose for the classification system developed by each subject was to provide a means for selecting a suitable gift, given a description of the person who was to receive it. The researchers found that (1) increased familiarity with the data resulted in a more precise classification scheme, (2) performance in retrieval
was improved when using one's own descriptors, and (3) most subjects used a hierarchical type structure, but all of them departed from this classification scheme at least once in designing their own retrieval system.

Durding, Becker, and Gould (6) conducted three data organization experiments designed to provide a quantitative view of how people organize data when they do not have to perform any subsequent tasks with the data. The subjects in these experiments were thirty-six paid undergraduates recruited from local liberal arts colleges. They were given word sets and asked to organize them on paper. The words in each set had a different predefined natural order or semantic relationship. For example, one of the word sets predefined as a network contained "tom", "bill", "beak", and "check". In this set "bill" has one semantic relationship with "tom", another with "beak", and a third with "check".

The first experiment was designed to determine whether the subjects would recognize the semantic relationships among the words and use them in the appropriate data structure. The researchers found that 79% of the hierarchical word sets were organized as hierarchies and 56% of the network word sets were organized as networks, but only 44% of the tabular word sets were organized into tables.

The next experiment provided a skeleton (spatially arranged sets of blanks connected by lines) of the data
organization to be used. The results of this experiment were as follows: (1) nearly all lists were correct, (2) 50 to 75% of the hierarchies and tables were correct, and (3) only 25% of the networks were correct.

In the last experiment the subjects were instructed to use a particular structure, but no skeleton was provided. The reported findings suggest that subjects were significantly faster at organizing words into the specific structures. The researchers also found that, when asked to use appropriate structures, 66% of the subjects were able to preserve the semantic relationships in the word sets. This was significantly better than in the first experiment where subjects were not asked to use a specific structure. During, Becker, and Gould believe that the task demands involved in these experiments were very closely related to what the computer system user would do when organizing data in his own mind for retrieval purposes.

Other data structure studies have found that (1) recall of categorized word lists was two to three times better with a hierarchical arrangement when compared to a random presentation (1); (2) when subjects were given clues in different structural arrangements and asked to solve "who done it" problems, they were inclined to change the mode of presentation for narrative and network descriptions (20); and (3) individuals were just as fast and accurate in responding to external test stimuli when using mental images
as they were when compared to actual perceptions (21).

The studies presented in this section use a variety of dependent variables to investigate the effectiveness of different data models. The reported findings indicate that subjects prefer the use of the hierarchical model (2), and also, perform best on non-query writing tasks (3) when using the hierarchical model. In contrast to this however, it is reported (17) that subjects perform best using the relational DML/data model when writing actual database queries. Another reported finding which is very relevant to this study, is that cognitive style as measured by the Klob Learning Style Inventory need not be considered in the selection of a database model (11).

These studies have contributed to increasing the understanding of the effect of different data models. However, the primary question--which data model provides the most effective external view of the database to support query writing by casual users? -- still remains unanswered.

2.1.2 Cognitive Style Studies

The problem of "user acceptance" in information systems, although by no means new, becomes more paramount as the number and diversity of end-users increase. During the late 1960s and early 1970s researchers began to recognize the effect of individual differences on the usability of computer-based information systems. Since this beginning, ever-increasing numbers of cognitive, personality, and
motivational studies have been conducted. Moran (18, p. 1) provides an introduction to "user psychology" and offers this as a new coherent subfield of computer science, in which the main contribution is to "reliably assure satisfactory user-computer interaction." Zmud (22, p. 966) states:

A decision maker's information requirements to a large extent are based upon the individual's "world-view," which is totally reflective of the individual. The objective of the aspect of MIS research on individual differences is to locate the critical individual differences and determine how best to design a MIS for individuals so characterized.

Much of the cognitive style research in MIS uses a variety of overlapping constructs, and generally the results have been inconsistent (13). The vast majority of the empirical studies on cognitive style--and more specifically on cognitive style as measured by the MBTI--has been done in fields other than MIS. Keen and Bronsema (13, p. 21) support the use of the MBTI in MIS research and identify two assumptions basic to cognitive style studies:

1. There are systematic differences among individuals in terms of perception, thinking, and judgment that significantly influences their choice of and response to information.

2. The difference between managers' and analysts' cognitive styles is a major explanation of difficulties in implementation.

Kilmann and Mitroff (14) conducted a study to determine whether individuals have different qualitative and quantitative notions of what constitutes an ideal
organization. In this experiment the researchers grouped managers by cognitive style (MBTI) and asked each to describe the ideal organization in his own words. They found strong similarities in the descriptions within individual personality groups and strong differences among the four groups. The researchers point out that the Sensing/Thinking (ST) and Intuition/Feeling (NF) groups, like the Intuition/Thinking (NT) and Sensing/Feeling (SF) groups, are extreme opposites. This, it is believed, "could have tremendous consequences for the design of modern large-scale organizations."

Another study, by Henderson and Nutt (10), involved executive MBAs and MHAs (Masters in Hospital Administration) from three large universities. Most subjects held upper level management positions in businesses and hospitals. The MBTI was administered, and each group was given capital investment cases designed for its particular personality type. Two dependent variables were measured: (1) the likelihood of approving the project and (2) the risk believed to be associated with the project. The reported results of this study suggest that a significant relationship existed between decision style and both the likelihood of project adoption and perception of risk. The researchers determined that "SF executives were most likely to adopt and saw the least risk, while ST executives were least likely to adopt and reported the most risk."
A study by Kaiser and Bostrom (12) tested the hypothesis that systems designers and users are different with respect to specific personality characteristics, as measured by the MBTI. In evaluating the results, with regard to the four main Jungian personality types, the researchers found a difference only in the SN dimension. A significantly higher percentage of S types compared to N types was found among users, whereas the designers were fairly balanced on this dimension. The findings in this study are totally inconsistent with the typical user/designer stereotypes portrayed in the literature, but the users in this study were members of the design team and not "typical" of the general user population.

Ghani (7) also found differences in the performance of Ts and Fs in a complex decision-making task while using reports of different formats. The F types preferred and performed better with graphic presentations; the Ts preferred and performed better with tabular presentations.

Based upon the reported findings in these studies it is anticipated that the student subjects used in this experiment will portray cognitive characteristics similar to the typical user stereotype. This would be reflected by a larger percentage of N types compared to S types and a larger percentage of F types compared to the number of T type individuals (12). Consistent with this, it seems reasonable to expect the Ns, Fs, and NFs to be more
confident and perform better using the relational model as compared to the same personality types using the other data models. The converse is expected for the opposite personality types. The Ss, Ts, and STs should be more confident and perform better using the network and hierarchical models compared to the relational data model (13, 7).

2.2 Needed Research

The empirical studies referenced in the previous sections have contributed significantly to the accumulation of knowledge about database query facilities in particular and information systems in general. However, much of the existing research was either not oriented towards casual user problems, or had design limitations that reduced its casual user applicability. These limitations, and a discussion of how this study will avoid them, are examined below:

(1) In other query system studies the subjects were not representative of the casual user population. Some of the studies used computer science majors who had completed a number of programming classes, others used professional computer programmers. The subjects to be used in this experiment will have little or no prior computer experience.

(2) The sample databases used in other studies were
over-simplified and did not present a realistic external view of a database within the business environment. The sales-related database to be introduced and used in this experiment provides sufficient complexity to evaluate the effect of the data model used in presenting the external view.

(3) Other studies have been designed to measure (a) users' preference of data models, (b) users' ability to memorize the contents of the database given different data models, (c) users' ability to illustrate and/or recognize the existing interrelationships given different data models, and (d) how users, given a choice, would organize the data themselves. This experiment will evaluate the user's ability to write a query for retrieving specific information given one of the three commonly used data models. The subjects will have to understand the contents of the database and the interrelationships among the entities to be able to write the queries required in the exercise.

(4) Other experiments have used commercially available query languages where the effect of the data model was confounded due to language complexities and other differences. To reduce the confounding effects of the differences in query language
difficulty upon the effect of the data model, only
one command will be used in the information
retrieval exercise. The one command will be the
same for each data model, and is designed to keep
the procedure for retrieving information as simple
as possible.

(5) This study will provide insights into the effect of
cognitive style differences as measured by the MBTI.
These different personality types will be evaluated
in terms of user performance and confidence in
writing database queries. The results of this study
can then be compared to Hoffer's (11) findings,
where the LSI was used to identify personality
differences.

In summary this study will provide insights into which
is the most effective data model by eliminating some of the
weaknesses of the studies discussed in Section 2.1.1. The
features of the study that will accomplish this objective
include: (1) the subjects used will be more representative
of the casual user population, (2) the database will be more
realistically representative of those found within a
business environment, (3) the experiment will involve
writing database queries, (4) one query language will be
used for all data models, eliminating the confounding
effects of the language and the data model, and (5) the
procedure for quantifying performance will involve evaluating different parts of the query.

The technique of evaluating individual parts of the query separately will permit more precise evaluation of the effect of the independent variable(s) by eliminating offsetting errors. For example, if only one score is used to represent performance, errors resulting from confusion because of the data model would be confounded by other errors, such as syntax errors that may be due to confusion resulting from characteristics of the DML.

2.3 Definition of Terms

Specific terms used in this study have the following definitions:

(1) Casual users - are users or potential users of database query facilities in a business environment who are not required as part of their job to use the system but who could benefit from the informational support (5). No prior experience with computers but familiarity with the data and application area are typical characteristics of this class of users.

(2) User confidence - is a dependent variable that will be self-reported. The measure will be reported on a scale of one to five. One indicates no confidence in the subject's own solution, and five
indicates total confidence.

(3) User performance - is a dependent variable measuring the correctness of the database queries written by the subjects in the experiment. This dependent variable is quantified by dividing the query into three parts and scoring each portion independently on a scale of zero to two. A detail description of the technique used to quantify performance is provided in Section 3.4.2.

(4) Cognitive style - refers to personality dimensions as measured by the MBTI.

(5) Data models - are the user's image (external view) of the data contained within the computing system.

(6) System usability - refers to casual user performance in constructing queries in English to gather specific information.

(7) Query facilities - refers to the contents of the database and the predefined procedure for retrieving information.
CHAPTER BIBLIOGRAPHY


3.1 Research Framework

The variable set being investigated in this study, and the makeup of these variables, is shown in Table I. The dependent variables included in this set were chosen as measures of database success. The independent variables are a subset of the variables presented in the descriptive model of the casual user's environment (Figure 1).

<table>
<thead>
<tr>
<th>INDEPENDENT VARIABLES</th>
<th>DEPENDENT VARIABLES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive Style</td>
<td>User Performance in Writing DB Queries</td>
</tr>
<tr>
<td>Sensing Types (S)</td>
<td>User Confidence in Writing DB Queries</td>
</tr>
<tr>
<td>Intuition Types (N)</td>
<td></td>
</tr>
<tr>
<td>Thinking Types (T)</td>
<td>Specification Portion</td>
</tr>
<tr>
<td>Feeling Types (F)</td>
<td>Condition Portion</td>
</tr>
<tr>
<td>ST Types</td>
<td>Navigation Portion</td>
</tr>
<tr>
<td>SF Types</td>
<td>Language Use</td>
</tr>
<tr>
<td>NT Types</td>
<td>Overall Performance</td>
</tr>
<tr>
<td>NF Types</td>
<td></td>
</tr>
<tr>
<td>Data Models</td>
<td></td>
</tr>
<tr>
<td>Relational</td>
<td></td>
</tr>
<tr>
<td>Hierarchical</td>
<td></td>
</tr>
<tr>
<td>Network</td>
<td></td>
</tr>
</tbody>
</table>

TABLE I
RESEARCH VARIABLE SET
The research hypotheses used in investigating the effects of the variables presented in Table I state that there are no interaction effects among the different levels of the independent variables. The model used to test this hypothesis compares each cell mean to the grand mean plus the column and row main effects. For example, if the user performance (UP) for say the network level of the data model factor (DM) and the sensing level of the cognitive style factor (CS) is determined by the additive effects of these two levels only, then no interaction is present. This would imply that the cell mean of the dependent variable intersecting at row i and column j could be written as (13):

$$\mu_{ij} = \mu_{..} + \alpha_i + \beta_j$$

for all i, j

where:

- $\mu_{ij}$ is the cell mean of the dependent variable intersecting at row i and column j,
- $\mu_{..}$ is the grand mean,
- $\alpha_i$ is the main effect of the ith level of factor A, computed as the difference between the mean of row i and the grand mean, and
- $\beta_j$ is the main effect of the jth level of factor B, computed as the difference between the mean of column j and the grand mean.

When the treatment means can be written in this form, it follows that the factors do not interact, or that the factor effects are additive. Interaction occurs when $\mu_{ij}$ can not be written in this form. To test for interaction effects
the following hypothesis would be examined:

\[ H_0: \mu_{ij} = \mu_{..} + \alpha_i + \beta_j \quad \text{for all } i, j \]

Given no interaction effects among the levels of the independent variables, each factor will be tested for main effects. To test for main effects in a two factor model, the following hypotheses would be examined (13):

**Factor A**

\[ H_0: \mu_1 = \mu_2 = \ldots = \mu_a \]

where \( \mu_1 \) is the mean of row 1.

**Factor B**

\[ H_0: \mu_1 = \mu_2 = \ldots = \mu_b \]

where \( \mu_1 \) is the mean of column 1.

The three hypotheses presented above will be tested for each of the five aspects of user performance (see Section 3.4.2), plus user confidence. This procedure of testing the hypotheses will be repeated for each of the three factorial designs presented in Section 3.5.

### 3.2 Research Hypotheses

To carry out the purpose of this study, three sets of hypotheses will be tested. Each hypothesis states that the effect of one or more independent variable will be the "same" on a given dependent variable. "Same" in these cases
refers to the mean of the dependent variable for each group involved in the test. Brief rationales and the resulting hypotheses are as follows:

(1) Studies have shown that people use mental images in problem-solving situations (1, 15, 16). These images are often highly individualized and may vary depending upon the problem context and cognitive characteristics of the individual (5). Although research is being conducted on data model equivalence and hybrid data models (3, 9, 18), no empirical evidence is available relating to the effect of all three commonly used data models on the casual user's performance in a query writing situation.

Hypotheses 1a and 1b are designed to investigate this area, and are based primarily on the research of Brosey and Shniederman (2). These researchers reported that using a hierarchical model with comprehension and memorization tasks resulted in better performance.

H1a - Casual user performance in writing database queries will be the same for each data model used.

H1b - Casual user confidence in writing database queries will be the same for each data model used.
(2) The second set of hypotheses are designed to test the effect of individual differences among cognitive style dimensions as measured by the MBTI. These hypotheses are intended to expand upon the Hoffer (5) study. Hoffer used the Klob Learning Style Inventory to identify cognitive style differences in an experiment designed to show data model preference when the user was not required to use a specific model. No empirical evidence is available relating to Jungian personality types' performance in a query writing situation.

Hypotheses 2a through 2f are designed to answer questions related to the effect of individual cognitive style differences in writing database queries.

H2a - Casual user performance in writing database queries will be the same for both levels of the Sensing/Intuition cognitive style dimension.

H2b - Casual user confidence in writing database queries will be the same for both levels of the Sensing/Intuition cognitive style dimension.

H2c - Casual user performance in writing database queries will be the same for both levels of the
Thinking/Feeling cognitive style dimension.

H2d - Casual user confidence in writing database queries will be the same for both levels of the Thinking/Feeling cognitive style dimension.

H2e - Casual user performance in writing database queries will be the same for all combinations of levels within the two cognitive style dimensions, Sensing/Intuition and Thinking/Feeling.

H2f - Casual user confidence in writing database queries will be the same for all combinations of levels within the two cognitive style dimensions, Sensing/Intuition and Thinking/Feeling.

(3) The third set of hypotheses, 3c through 3f, are designed to answer questions related to the interaction effects of data models and cognitive style classifications on user performance and confidence in writing database queries.

H3a - The effect of a specific external data model upon user performance will be the same for both levels of the Sensing/Intuition cognitive style dimension.
H3b - The effect of a specific external data model upon user confidence will be the same for both levels of the Sensing/Intuition cognitive style dimension.

H3c - The effect of a specific external data model upon user performance will be the same for both levels of the Thinking/Feeling cognitive style dimension.

H3d - The effect of a specific external data model upon user confidence will be the same for both levels of the Thinking/Feeling cognitive style dimension.

H3e - The effect of a specific external data model upon user performance is the same for all combinations of levels within the two cognitive style dimensions, Sensing/Intuition and Thinking/Feeling.

H3f - The effect of a specific external data model upon user confidence is the same for all combinations of the two cognitive style dimensions, Sensing/Intuition and Thinking/Feeling.
3.3 Research Design

This study will be a laboratory experiment conducted in the classroom using three experimental groups, one for each data model. The data models will be used to present three different external views for the same data entities.

The experiment will be presented to the students as a classroom exercise. Three sections of a graduate level class at North Texas State University will be used. The class, entitled "Computer Based Systems", is an introductory course, required for graduate level, non-MIS majors having no previous computer course work. Because of their work experience, education, and lack of computer experience, these students provide a sample that is representative of the population of business professionals having little or no computing background.

The study will examine the independent effects and the interaction effects of two variables (cognitive style and data models) upon the two dependent variables (casual user's performance and confidence). Among the available tested instruments designed to measure a variety of cognitive dimensions (4, 7, 11, 17) the MBTI measuring Jungian typology was chosen for three reasons. First, it has been used in a number of MIS research studies because of the similarities between the Jungian classifications and managerial styles (8, 10). The second reason for the selection of the MBTI is the increasing body of empirical
knowledge being accumulated with respect to these
personality types (7, 20). The third reason is to extend
the Hoffer study, which used the Kolb Learning Style
Inventory (LSI). Hoffer (5, p. 165) says, "further
research is needed to address the relationships between
database architecture and cognitive variables beyond those
covered in this study."

The MBTI measures four Jungian dimensions: (1)
Extroversion/Introversion (EI), (2) Sensing/Intuition (SN),
(3) Thinking/Feeling (TF), and (4) Judging/Perceptive (JP).
All of the dimensions should be viewed as continua, in which
individuals are classified as having tendencies toward one
extreme or another. It is not suggested that any
classification is superior to another but, rather, that
people simply have one personality tendency or the other (6,
7). The four dimensions are indications of the following:
(1) the E/I dimension suggests that people relate to the
outer world of people and things (E) versus the inner world
of ideas (I); (2) the S/N dimension suggests that a person
would rather work with known facts (S) versus looking for
possibilities (N); (3) the T/F dimension suggests that a
person bases judgments more on impersonal analysis and logic
(T) versus personal values (F); and (4) the J/P dimension
suggests that a person would prefer a planned, orderly way
of life (J) versus a flexible, spontaneous one (P).

Because of their relationship to information systems
activities, this study is concerned with the second and third of these dimensions (10). The second dimension is an indication of how people "take in information," and the third dimension is an indication of how people "use information" in decision making. Figure 2 presents a model of these two dimensions and the indicated personality traits.

The MBTI is a tested instrument for identifying personality types, and has been used for a number of years in psychology and educational research (12). In more recent years, the instrument has become increasingly popular as a tool for identifying differences among individuals in business settings for research dealing with human factors (7, 10, 11).
Thinking (T) - relies on cognitive processes; judgments based on logic.

Sensing (S) - relies on known facts; attention to detail.

Intuition (N) - sees objects as possibilities; focuses on concepts.

Feeling (F) - relies on affective processes; judgments based on feelings.

* Horizontal dimension represents the way people gather information.

** Vertical dimension represents the way people process information.

FIGURE 2 - MODEL OF JUNGIAN PERSONALITY TYPES
3.4 Procedures for Collection of Data

The data collection process will be divided into two phases. Phase I will be the administration of the MBTI to all subjects. The data collected from the administration of this instrument will be used to classify the subjects into four Jungian psychological types: (1) Sensing/Feeling, (2) Intuition/Feeling, (3) Sensing/Thinking, and (4) Intuition/Thinking. The personality types will be established prior to assigning the subjects to one of the three experimental groups. The purpose for this is to insure an equal representation of personality types in each experimental group. The subjects from each personality group will then be randomly selected and alternately assigned to one of the three experimental groups, hereafter referred to as Groups A, B, and C.

Phase II of the data collection process will be the administration of an information retrieval exercise. The exercise includes a test designed to permit the evaluation of the subject's ability to comprehend and retrieve information from a database by writing English-like queries. Section 3.4.1 contains a detailed description of the contents of the exercise.

The exercises to be used in Phase II differ only in the data model used in presenting the external view of the database and the narrative description of this external view. Group A will receive exercises using the Relational
Model, Group B will receive the Hierarchical Model, and Group C will receive the Network Model. The exercise describes a business situation for which information is needed to make decisions. The subjects will be asked to write queries to retrieve data presented in the external view of the database. In addition to posing queries to accomplish the information retrieval, the subjects will also be requested to indicate their confidence (on a scale of one to five) that their instructions would in fact provide the information needed.

The experimental exercise will be made an integral and graded part of the learning experience in the Computer Based Systems course. The course introduces students to computers and computer based information systems. One component of the course is an introduction to Database Management Systems and systems architecture. The experimental exercise will be used as a vehicle for getting students involved and interested in this topic, but will be administered prior to the introduction of the database topic.

Each student will be given two course grades for completing the exercise. The purpose for these grades is to provide incentive for students to participate conscientiously. The grades will be equivalent in weight to those assigned in other classroom exercises. The first grade will be a measurement of the accuracy of the queries in providing the desired information, and the second will be
based on the correlation between the confidence level indicated by the subject and the accuracy of the query for each problem. For example, if the query posed by the subject does provide adequate information but his confidence level is low, the second grade will be low (equivalent to the level of lack of confidence). This second grade is designed to encourage the subjects to conscientiously assess the accuracy of their response in each problem-solving situation, and to report this confidence level.

3.4.1 Exercises Used

The exercises to be used in this laboratory experiment are designed as self-contained instruments to be administered in one two hour period. There are a total of four booklets. Examples of these booklets are presented in appendices A, B, C, and D. The first three booklets, entitled "Exercise Booklet A", "B" and "C", differ only in the data model used to present the external view. Exercise Booklet A presents the relational model, Exercise B the hierarchical model, and Exercise Booklet C presents the network data model. The fourth booklet, titled "Information Retrieval Problems", will be the same for each of the three experimental groups. Each subject will receive only one of the three exercise booklets, and all subjects will receive booklet four.

The exercise booklets (A, B, and C) are divided into four parts. Parts I, III, and IV are identical for each of
the three exercises. Part I introduces the subject to the experiment, the hypothetical organization for which the database was designed, and four major entities belonging to this organization. These entities include "Customers", "Customer Orders", "Finished-Goods Inventory", and "Parts Inventory", and are presented in a Management View in Part I. Part II, which is different for each of the three exercises, provides a detailed description of the four entities introduced in Part I. This description includes (1) illustrations of both the external view and a sample of the database, (2) a narrative description of each attribute defined into the database, and (3) a description with examples of how the entities are interrelated. Part III of each exercise booklet is used to define the procedure which must be used to retrieve information. This includes (1) a description of the general requirements and definitions of terminology used, (2) a description of the information retrieval command, and (3) examples of how this command may be used. Part IV provides two practice problems. Solutions to the practice problems are presented on the last page of the exercise booklet to give the subjects feedback on their understanding and use of the information retrieval command.

The exercises used in this experiment had the following design criteria:

1 - The data entities and interrelationships defined in the sample database and presented in the external view were
to be representative of a typical business organization. These entities and interrelationships were to be used in the evaluating the subject's ability to comprehend the database and retrieve information, given a specific data model.

2 - The experiment was to measure the effects of the data model on the subject's ability to write database queries. These effects were not to be biased by the effects of the query language. To meet this objective the query language design must (a) be the same for each data model group, (b) be easy to learn, and (c) require the subject to state explicitly the navigational path used in retrieving the desired information. By requiring the navigational path to be explicitly stated, it can be determined if the subject does in fact comprehend the interrelationships among the entities involved in the information retrieval.

3 - In an effort to retain internal validity by eliminating subject interaction, the experiment is to be conducted in one period. To accomplish this objective without over simplifying the database, the following design characteristics were built into the experiment:
(a) The number of attributes per entity was limited. This characteristic reduced the length of the narrative description of the database, the size of the
illustrations, and the time required for the subjects to study and comprehend the contents.

(b) The query language is limited to one English-like retrieval command. This command identifies the attributes to be retrieved and their location in the database, uses a "when" clause for record or tuple selection, and a "thru" clause to define the desired navigational path among the entities.

4 - The subject's performance in writing the queries is not to be confounded by variations in problem complexity, or carry over effects resulting from solving progressively more complex problems. To meet this objective the exercise booklets are designed as tutorials. Each of which is climaxed with two practice problems having increasing levels of difficulty. The subjects will then be given feedback on the correctness of their solutions to the practice problems through the use of sample solutions.

The three problems presented in booklet four are of approximately the same level of difficulty. The combined score on these three problems will be used as a measure of the subject's performance.

3.4.2 Grading the Exercises

The technique to be used in quantifying the subject's performance in writing database queries will be different
from techniques used in other database studies (14, 19). In these studies, each query was assigned to a classification which represented its correctness. This technique did not permit the isolation of specific types of errors, the evaluation of difficulties with specific aspects of the query, or the analysis of multiple error types within the same query. For example, in the Welty (19) experiment, which was designed to measure the effect of procedurality by comparing the performance of subjects using SQL to the performance of subjects using TABLET, classified the queries written by each subject into one of the following:

1 - Completely correct (CC) -- No errors of any kind.

2 - Minor language error (ML) -- The solution was basically correct but had a small error that would be found by a reasonably good translator.

3 - Minor operand error (MO) -- The solution has a minor error in its data specification, perhaps a misspelled column name.

4 - Minor substance error (MS) -- The solution yields a result that is not quite correct but its incorrectness is due to the statement of the problem.

5 - Correctable (CO) -- The solution is wrong but correctable by a good compiler.

6 - Major substance error (XS) -- The query is syntactically correct but answers a different question than the one specified.

7 - Major language error (XF) -- A major error in the syntax (form) of the language has been made.

8 - Incomplete (IN) -- The query was not finished.

9 - Unattempted (UN) -- No solution was attempted.
The query language being used in this study is limited to one retrieval command. The purpose for this is to reduce the confounding effects of the difficulty of the query language upon the effect of the data model being used. This "LIST" command consists of (1) a specification portion, where the user must specify all attributes to be retrieved from the database, along with record types or relations, (2) a condition portion (optional), where the user must identify the condition(s), if any, under which the attributes are to be retrieved, and (3) a navigation portion (also optional), where the user must define the navigational path to be used in the information retrieval process. The optional navigation clause must be used when multiple relations or record types are involved in the information retrieval request. The specific conditions requiring the use of this clause are defined in the exercise booklets and vary slightly for each data model. All three problems in the fourth booklet (Appendix D) require the use of both optional clauses. The following example illustrates the three parts of the "LIST" command used in this study:

Specification portion - LIST PART-DESC in PART-INV
Condition portion - when FG-UNIT-PR in FG-INV < 50.00,
Navigation portion - THRU MAT-EXP.

Each of the three parts of the queries will be evaluated separately and received a score of zero, one, or
two. A grade of zero means that portion of the query is correct, one means that portion of the query has a minor error, and two means that portion has a more serious error. Errors in the use of the language will not be considered when grading the individual parts of the query. Each query will receive a fourth grade of zero, one, or two that represents the correctness in using the language or command.

After writing each of the three queries in the fourth booklet, the subjects will be asked to indicate their confidence in the correctness of the query just written. Their confidence is reported by circling a number from one to five, indicating low to high user confidence.

All queries will be graded independently by two graders, each of which will prepare grade sheets reflecting the subject's confidence rating, specification score, condition score, navigation score, and language score for each of the three problems. Figure 3 provides an example of the grade sheets to be used in this experiment.

Consistency in grading the three queries developed by each subject will be maintained in the following manner. Each grader will evaluate the solution to problem one for all subjects and post these scores to his set of grade sheets. This technique will be repeated for problem number two, and then again for problem three. All variations between the two sets of grade sheets will be reviewed and resolved by the graders. This process will involve the re-
examination by both graders of all queries in question. The difference in the original scores will then be resolved.

The three sets of scores for each subject will then be summarized on the grade sheets. The summarization will contain the subject's average confidence (ranging from 1 - 5), a sum for each of the other classifications (ranging from 0 - 6), and an overall performance score, which will be the sum of the total scores for each of the four areas. This sum of the total scores, which represents the subjects overall performance ranges from 0 to 24. Each portion of the performance score will then be evaluated separately using two-way Analysis of Variance.

<table>
<thead>
<tr>
<th></th>
<th>Confidence</th>
<th>Specifications</th>
<th>Condition</th>
<th>Navigation</th>
<th>Syntactical</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROBLEM 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROBLEM 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROBLEM 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3 - Sample Grade Sheets
Section 3.5 presents a detail description of the experimental models used in the factorial design and the planned statistical analysis.

Performance is broken down into four scores for each query because one of the main objectives of this research is to determine the effect of the data model on a casual user's performance in writing database queries. It is believed that the comprehension of interrelationships among entities and the corresponding navigational specifications in query writing may be influenced more by the data model than the other three aspects of writing database queries. If only one score were assigned for each query, it would have to reflect overall correctness. The use of a single performance score would unduely aggregate the effects of potentially significant aspects of user performance.

3.5 Procedures for Analyzing the Data

The data from this experiment will be analyzed using Two-way Analysis of Variance (ANOVA) to test for a statistically significant difference in user performance and confidence among the three experimental groups. The two dependent variables (user performance [all aspects] and confidence) will be analyzed separately in the univariate design. This technique was chosen to permit both the evaluation of main effects and the evaluation of the interaction effects of the two independent variables (data models and Jungian Personality Types).
For analysis purposes, the personality type is defined as three factors. Two of the three factors have two levels each, while the third factor is divided into four levels. The first personality factor (S/N dimension) is an indication of how subjects take in information and is made up of the two levels, Sensing (S) and Intuition (N). The second personality factor (T/F dimension), indicating the way people process information, consists of the Thinking (T) level and the Feeling (F) level. In addition, all combinations of these two Jungian personality dimensions will be evaluated. This third factor consists of four levels which include: (1) the ST types, (2) the NT types, (3) the SF types, and (4) the NF types.

These design decisions require the use of three experimental models in the factorial design. The three models will permit the evaluation of the two Jungian dimensions from an independent standpoint as well as the combined effects. In this design, Factor A in each experimental model represents the data models used in the experiment, and consists of three levels: A1 = relational data model, A2 = hierarchical model, and A3 = network model.

In experimental model 1 (Figure 4), Factor B represents the way people take in information: B1 = Sensing types and B2 = Intuition types. This experimental model will be used in testing Hypotheses 1a, 1b, 2a, 2b, 3a, and 3b.
In Experimental Model 2 (Figure 5), Factor B represents the way information is used by individuals: B1 = Thinking types, and B2 = Feeling types. This model will be used in testing Hypotheses 2c, 2d, 3c, and 3d.
In Experimental Model 3 (Figure 6), Factor B represents all combinations of these two Jungian dimensions: B1 = the Sensing/Thinking (ST) types, B2 = the Intuition/Thinking (NT) types, B3 = the Sensing/Feeling (SF) types, and B4 = the Intuition/Feeling (NF) types. This model will be used in testing Hypotheses 2e, 2f, 3e, and 3f.
FACTOR A - DATA MODELS  FACTOR B - PERSONALITY TYPES

A1 = Relational  B1 = Sensing/Thinking Types (ST)
A2 = Hierarchical  B2 = Intuition/Thinking Types (NT)
A3 = Network  B3 = Sensing/Feeling Types (SF)
          B4 = Intuition/Feeling Types (NF)

Figure 6 - Factorial Design Model 3

Table II provides a summary of how the three experimental models will be used to test the fourteen hypotheses, along with the dependent and independent variables involved.
<table>
<thead>
<tr>
<th>HYP</th>
<th>EXP.</th>
<th>DEPENDENT VARIABLE</th>
<th>TESTING FOR THE EFFECT OF</th>
</tr>
</thead>
<tbody>
<tr>
<td>la</td>
<td>1</td>
<td>Performance</td>
<td>Data Model Used</td>
</tr>
<tr>
<td>lb</td>
<td>1</td>
<td>Confidence</td>
<td>Data Model Used</td>
</tr>
<tr>
<td>2a</td>
<td>1</td>
<td>Performance</td>
<td>Sensing/Intuition Level</td>
</tr>
<tr>
<td>2b</td>
<td>1</td>
<td>Confidence</td>
<td>Sensing/Intuition Level</td>
</tr>
<tr>
<td>2c</td>
<td>2</td>
<td>Performance</td>
<td>Thinking/Feeling Level</td>
</tr>
<tr>
<td>2d</td>
<td>2</td>
<td>Confidence</td>
<td>Thinking/Feeling Level</td>
</tr>
<tr>
<td>2e</td>
<td>3</td>
<td>Performance</td>
<td>Combinations of S/N and T/F</td>
</tr>
<tr>
<td>2f</td>
<td>3</td>
<td>Confidence</td>
<td>Combinations of S/N and T/F</td>
</tr>
<tr>
<td>3a</td>
<td>1</td>
<td>Performance</td>
<td>Interaction of DM and S/N</td>
</tr>
<tr>
<td>3b</td>
<td>1</td>
<td>Confidence</td>
<td>Interaction of DM and S/N</td>
</tr>
<tr>
<td>3c</td>
<td>2</td>
<td>Performance</td>
<td>Interaction of DM and T/F</td>
</tr>
<tr>
<td>3d</td>
<td>2</td>
<td>Confidence</td>
<td>Interaction of DM and T/F</td>
</tr>
<tr>
<td>3e</td>
<td>3</td>
<td>Performance</td>
<td>Interaction of DM and Comb.</td>
</tr>
<tr>
<td>3f</td>
<td>3</td>
<td>Confidence</td>
<td>Interaction of DM and Comb.</td>
</tr>
</tbody>
</table>
3.6 **Limitations**

The following deficiencies in the experimental process will affect the ability to generalize the conclusions of this laboratory experiment:

1. The student subjects used in this experiment may be less knowledgeable about the data entities and the relationships among these entities than managers in a business application area.

2. No provisions, other than random selection for the experimental groups, have been made to control for the effect of variables such as learning skills and self-expression.

3. The dependent variable 'user confidence' is a self-reported value on a scale of one to five. As reported in section 3.4 efforts were made to encourage the subjects to conscientiously record their confidence in the accuracy of each query written. However, it is recognized that any significant differences in confidence among the Jungian types may be due in part to innate personality differences, and not the effects of the experiment.

4. The material learned by the subjects in this experiment is presented as a tutorial in narrative form. No provisions have been made to control for
other learning techniques that might be preferred by the different personality types.

The findings of this study are valid only for the population studied. However, previous studies in this area have used similar populations, which will allow for meaningful comparisons of the results. Care must be taken in the extension of any conclusions drawn from this study to the general population of casual users.
CHAPTER BIBLIOGRAPHY


CHAPTER IV
Results and Analysis

4.1 Analysis of Data

The MBTI answer sheets collected from Phase I of this study were scored, and the Jungian personality classifications were established by the Testing and Counseling Center at North Texas State University. The queries posed by each subject in Phase II of the experiment were graded independently by two trained graders. All differences between the two evaluations were resolved by the graders. Means were then computed for each subject's confidence, and sums were accumulated for the four error classifications for each subject. The sums representing the subject's performance in writing a specific portion of the query were then totaled to represent the subject's overall performance.

These data were then used with the Analysis of Variance to test the null hypothesis that no difference exists among the means as stated for each research hypothesis. All main effects were further analyzed using Tukey's test for comparing all possible pairs of group means for that factor. The significance below which each hypothesis is to be rejected was set at the .05 level.
4.2 Subject Demographics

The sixty-three subjects used in this experiment were MBA students at North Texas State University. All were enrolled in Computer Based Systems (an introduction to computer based information systems) for the Spring Semester, 1984. This group of students was selected because of their business backgrounds and their limited familiarity with computers and computer programming.

The following demographic information was collected using the information sheets presented in Appendix E: (1) the ages of the subjects ranged from twenty-one to fifty-six with a mean of twenty-nine and a standard deviation of 6.5; (2) the group was made up of thirty males and thirty-three females; (3) work experience in a business environment ranged from zero to twenty years with a mean of three years (67% reported at least one year of experience); (4) for the number of computer-related classes completed, 71% reported zero, 11% reported one, 7% reported two, with the remaining 11% reporting from three to five classes; (5) for the number of months of work experience as a direct user of a computer system, 76% reported zero, 8% reported less than three months, and the remaining 16% was spread over a range of five to forty-eight months; (6) for the number of months of programming experience in a business environment, 89% reported zero, and the remaining 11% reported from two to thirty-six months experience.
These subjects as a group have characteristics similar to those reported for the casual user population. For example, the demographic information collected does reflect limited experience with computer systems and a reasonably strong business background.

4.3 Jungian Classifications

The Jungian classifications resulting from the administration of the MBTI in Phase II of this experiment were established. The subjects were then divided into groups based upon the personality type. From these groups the subjects were then randomly selected and alternately assigned to one of the three data model groups. This assignment resulted in the distributions presented in Tables III, IV, and V.

### TABLE III

**CELL FREQUENCIES FOR THE SENSING-INTUITION DIMENSION BY DATA MODELS**

<table>
<thead>
<tr>
<th>DATA MODELS</th>
<th>SENSING</th>
<th>INTUITION</th>
<th>TOTALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relational</td>
<td>17</td>
<td>6</td>
<td>23</td>
</tr>
<tr>
<td>Hierarchical</td>
<td>15</td>
<td>7</td>
<td>22</td>
</tr>
<tr>
<td>Network</td>
<td>13</td>
<td>5</td>
<td>18</td>
</tr>
<tr>
<td>Totals</td>
<td>45</td>
<td>18</td>
<td>63</td>
</tr>
</tbody>
</table>
### TABLE IV

**CELL FREQUENCIES FOR THE THINKING-FEELING DIMENSION BY DATA MODELS**

<table>
<thead>
<tr>
<th>DATA MODELS</th>
<th>THINKING</th>
<th>FEELING</th>
<th>TOTALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relational</td>
<td>17</td>
<td>6</td>
<td>23</td>
</tr>
<tr>
<td>Hierarchical</td>
<td>15</td>
<td>7</td>
<td>22</td>
</tr>
<tr>
<td>Network</td>
<td>11</td>
<td>7</td>
<td>18</td>
</tr>
<tr>
<td>Totals</td>
<td>43</td>
<td>20</td>
<td>63</td>
</tr>
</tbody>
</table>

### TABLE V

**CELL FREQUENCIES FOR JUNGIAN PERSONALITY TYPES BY DATA MODELS**

<table>
<thead>
<tr>
<th>DATA MODELS</th>
<th>ST'S</th>
<th>NT'S</th>
<th>SF'S</th>
<th>NF'S</th>
<th>TOTALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relational</td>
<td>13</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>23</td>
</tr>
<tr>
<td>Hierarchical</td>
<td>11</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td>Network</td>
<td>9</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td>Totals</td>
<td>33</td>
<td>10</td>
<td>12</td>
<td>8</td>
<td>63</td>
</tr>
</tbody>
</table>

#### 4.4 Descriptive Statistics

As was discussed in Section 3.4.2 (Grading the Exercises), each of the three queries written by the subjects was assigned four grades. These grades are integer values ranging from zero to two. Each grade represents the
subject's performance in regard to a specific aspect of the query written. A grade of zero means there were no errors in that portion of the query, a one indicates the subject made a minor error on that portion, and a two means a more serious error was made. The four aspects of each query being evaluated were (1) the specification portion of the query, (2) the condition portion, (3) the navigation portion, and (4) the use of the language (syntactical).

Because each subject was required to write three queries, the performance for a given aspect ranges from zero to six. A score of six means that there was a serious error in this portion on all three queries written by the subject. The mean performance of all subjects, for each of the four parts of the query, and the percentage of subjects making no errors are presented in Table VI.
TABLE VI

MEAN PERFORMANCE SCORE (RANGE 0 - 6) AND PERCENTAGES OF SUBJECTS MAKING NO ERRORS

<table>
<thead>
<tr>
<th>ERROR TYPES</th>
<th>MEAN PERFORMANCE</th>
<th>PERCENT HAVING NO ERRORS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specification</td>
<td>1.4</td>
<td>60.3</td>
</tr>
<tr>
<td>Condition</td>
<td>2.2</td>
<td>22.2</td>
</tr>
<tr>
<td>Navigation</td>
<td>2.8</td>
<td>12.7</td>
</tr>
<tr>
<td>Language Use</td>
<td>2.7</td>
<td>15.9</td>
</tr>
</tbody>
</table>

The information in this table illustrates that the subjects had less trouble with the specification portion of the query than with the other two parts. This is not surprising because the English statement of each problem stated explicitly what information was to be retrieved. However, to write the condition portion of the query, the subject had to understand what the English statement of the problem was asking for, and then translate this into an acceptable form when writing the query. To write the navigation portion, the subjects had to identify where within the database the information being retrieved was located, and then specify the linkage among these entities.

Tables VII - X present frequency distributions of performance scores broken down into each of the four aspects.
TABLE VII
FREQUENCY DISTRIBUTION OF SPECIFICATION PERFORMANCE SCORES

<table>
<thead>
<tr>
<th>SCORE</th>
<th>FREQ.</th>
<th>PERCENT</th>
<th>CUMM. PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>38</td>
<td>60.3</td>
<td>60.3</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>6.3</td>
<td>66.7</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>7.9</td>
<td>74.6</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>6.3</td>
<td>81.0</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>6.3</td>
<td>87.3</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>12.7</td>
<td>100.0</td>
</tr>
</tbody>
</table>

63 100.0

TABLE VIII
FREQUENCY DISTRIBUTION OF CONDITION PERFORMANCE SCORES

<table>
<thead>
<tr>
<th>SCORE</th>
<th>FREQ.</th>
<th>PERCENT</th>
<th>CUMM. PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>14</td>
<td>22.2</td>
<td>22.2</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>3.2</td>
<td>34.9</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
<td>34.9</td>
<td>69.8</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>7.9</td>
<td>77.8</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>7.9</td>
<td>85.7</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>14.3</td>
<td>100.0</td>
</tr>
</tbody>
</table>

63 100.0
TABLE IX
FREQUENCY DISTRIBUTION OF NAVIGATION PERFORMANCE SCORES

<table>
<thead>
<tr>
<th>SCORE</th>
<th>FREQ.</th>
<th>PERCENT</th>
<th>CUMM. PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8</td>
<td>12.7</td>
<td>12.7</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3.2</td>
<td>15.9</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>47.6</td>
<td>63.5</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1.5</td>
<td>65.1</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>12.7</td>
<td>77.8</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>7.9</td>
<td>85.7</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>14.3</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>63</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100.0</td>
</tr>
</tbody>
</table>

TABLE X
FREQUENCY DISTRIBUTION OF LANGUAGE USE PERFORMANCE SCORES

<table>
<thead>
<tr>
<th>SCORE</th>
<th>FREQ.</th>
<th>PERCENT</th>
<th>CUMM. PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10</td>
<td>15.9</td>
<td>15.9</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>7.9</td>
<td>23.8</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>25.4</td>
<td>49.2</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>23.8</td>
<td>73.0</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>9.5</td>
<td>82.5</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>7.9</td>
<td>90.5</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>9.5</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>63</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100.0</td>
</tr>
</tbody>
</table>
A score representing the subject's overall performance was computed as the sum of the four error type scores. Consequently the subject's overall performance scores will range from zero to twenty-four. The overall mean performance score was 9.1, with 4.8% making no errors. A frequency distribution of the overall performance scores is presented in Table XI.

**TABLE XI**

**FREQUENCY DISTRIBUTION OF OVERALL PERFORMANCE SCORES**

<table>
<thead>
<tr>
<th>SCORE</th>
<th>FREQ.</th>
<th>PERCENT</th>
<th>CUMM. PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-2</td>
<td>7</td>
<td>11.1</td>
<td>11.1</td>
</tr>
<tr>
<td>3-5</td>
<td>13</td>
<td>20.6</td>
<td>31.7</td>
</tr>
<tr>
<td>6-8</td>
<td>15</td>
<td>23.8</td>
<td>55.5</td>
</tr>
<tr>
<td>9-11</td>
<td>10</td>
<td>15.9</td>
<td>71.4</td>
</tr>
<tr>
<td>12-14</td>
<td>7</td>
<td>11.1</td>
<td>82.5</td>
</tr>
<tr>
<td>15-17</td>
<td>3</td>
<td>4.8</td>
<td>87.3</td>
</tr>
<tr>
<td>18-20</td>
<td>2</td>
<td>3.2</td>
<td>90.5</td>
</tr>
<tr>
<td>21-24</td>
<td>6</td>
<td>9.5</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>63</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

The subjects' confidence level is a self-reported value for each of the three queries, on a scale of one to five. One reflects no confidence in the user's own solution to the specific problem, and five reflects total confidence. The average confidence for the three queries was computed for
each subject. The mean confidence for all subjects was 3.9 (see Table XII for a distribution of these frequencies).

TABLE XII

FREQUENCY DISTRIBUTION OF SUBJECTS' CONFIDENCE LEVEL

<table>
<thead>
<tr>
<th>CONFIDENCE</th>
<th>FREQ.</th>
<th>PERCENT</th>
<th>CUMM. PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>High - 5</td>
<td>21</td>
<td>33.3</td>
<td>33.3</td>
</tr>
<tr>
<td>4</td>
<td>22</td>
<td>34.9</td>
<td>68.2</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>22.2</td>
<td>90.4</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4.8</td>
<td>95.2</td>
</tr>
<tr>
<td>Low - 1</td>
<td>3</td>
<td>4.8</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>63</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

4.5 Statistical Analysis

The data from this experiment was analyzed using Two-way Analysis of Variance (ANOVA) to test for a statistically significant difference in user performance and confidence among the three experimental groups. The two dependent variables (user performance [all aspects] and confidence) were analyzed separately in the univariate design. This technique was chosen to permit both the evaluation of main effects and the evaluation of the interaction effects of the two independent variables (data models and Jungian Personality Types).

For analysis purposes, the personality type was defined
as three factors. Two of the three factors have two levels each, while the third factor is divided into four levels. The first personality factor (S/N dimension) is an indication of how subjects take in information and is made up of the two levels, Sensing (S) and Intuition (N). The second personality factor (T/F dimension), indicating the way people process information, consists of the Thinking (T) level and the Feeling (F) level. In addition, all combinations of these two Jungian personality dimensions are evaluated. This third factor consists of four levels which include: (1) the ST types, (2) the NT types, (3) the SF types, and (4) the NF types. The statistical analysis involved the use of the three experimental models presented and discussed in Section 3.5. Figure 7 through Figure 11 are histograms presenting the ANOVA cell means for all values tested in the experimental models. Figure 13 is a tabular presentation of the ANOVA cell means and the corresponding p values.
Figure 7 - Specification Errors

Figure 8 - Condition Errors
Figure 9 - Navigation Errors

Figure 10 - Language Use Errors
Figure 11 - Overall Performance

Figure 12 - USLR Confidence
<table>
<thead>
<tr>
<th></th>
<th>CONF.</th>
<th>SPEC.</th>
<th>COND.</th>
<th>NAVIG.</th>
<th>SYN.</th>
<th>O. P.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relational</td>
<td>4.0</td>
<td>.8</td>
<td>1.7</td>
<td>2.0</td>
<td>2.7</td>
<td>7.2</td>
</tr>
<tr>
<td>Hierarchical</td>
<td>3.5</td>
<td>2.2</td>
<td>2.6</td>
<td>3.6</td>
<td>2.7</td>
<td>11.1</td>
</tr>
<tr>
<td>Network</td>
<td>4.2</td>
<td>1.3</td>
<td>2.5</td>
<td>2.8</td>
<td>2.5</td>
<td>9.1</td>
</tr>
<tr>
<td>p</td>
<td>(.18)</td>
<td>(.05)</td>
<td>(.30)</td>
<td>(.01)</td>
<td>(.92)</td>
<td>(.10)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>CONF.</th>
<th>SPEC.</th>
<th>COND.</th>
<th>NAVIG.</th>
<th>SYN.</th>
<th>O. P.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensing</td>
<td>3.8</td>
<td>1.7</td>
<td>2.4</td>
<td>2.8</td>
<td>3.0</td>
<td>9.9</td>
</tr>
<tr>
<td>Intuition</td>
<td>4.1</td>
<td>.7</td>
<td>1.9</td>
<td>2.8</td>
<td>1.8</td>
<td>7.2</td>
</tr>
<tr>
<td>p</td>
<td>(.40)</td>
<td>(.09)</td>
<td>(.36)</td>
<td>(.96)</td>
<td>(.02)</td>
<td>(.12)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>CONF.</th>
<th>SPEC.</th>
<th>COND.</th>
<th>NAVIG.</th>
<th>SYN.</th>
<th>O. P.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thinking</td>
<td>4.1</td>
<td>1.3</td>
<td>2.1</td>
<td>2.6</td>
<td>2.6</td>
<td>8.6</td>
</tr>
<tr>
<td>Feeling</td>
<td>3.4</td>
<td>1.6</td>
<td>2.5</td>
<td>3.3</td>
<td>2.8</td>
<td>10.2</td>
</tr>
<tr>
<td>p</td>
<td>(.01)</td>
<td>(.67)</td>
<td>(.38)</td>
<td>(.14)</td>
<td>(.76)</td>
<td>(.34)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>CONF.</th>
<th>SPEC.</th>
<th>COND.</th>
<th>NAVIG.</th>
<th>SYN.</th>
<th>O. P.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST Types</td>
<td>4.0</td>
<td>1.5</td>
<td>2.3</td>
<td>2.6</td>
<td>2.9</td>
<td>9.2</td>
</tr>
<tr>
<td>NT Types</td>
<td>4.5</td>
<td>.9</td>
<td>1.5</td>
<td>2.4</td>
<td>1.7</td>
<td>6.5</td>
</tr>
<tr>
<td>SF Types</td>
<td>3.3</td>
<td>2.3</td>
<td>2.7</td>
<td>3.3</td>
<td>3.2</td>
<td>11.6</td>
</tr>
<tr>
<td>NF Types</td>
<td>3.5</td>
<td>.5</td>
<td>2.4</td>
<td>3.2</td>
<td>2.0</td>
<td>8.1</td>
</tr>
<tr>
<td>p</td>
<td>(.03)</td>
<td>(.23)</td>
<td>(.55)</td>
<td>(.53)</td>
<td>(.12)</td>
<td>(.27)</td>
</tr>
</tbody>
</table>

Figure 13 -- ANOVA Cell Means and Corresponding p Values
4.6 Testing the Hypotheses

Hypotheses 1a and 1b are designed to evaluate the effect of the three commonly used data models on casual user's performance and confidence in a query writing exercise. Hypothesis 1a states: "Casual user performance in writing database queries will be the same for each data model used." As stated earlier, the word "same" as used in each hypothesis, refers to the cell mean of the dependent variable for each level in the experimental model.

In evaluating the five performance aspects using Experimental Model 1 (Figure 4), the results of the ANOVA tests indicate a significant main effect, due to the data model used, on the subject's performance in writing both the specification portion and the navigation portion of the queries. Further analysis using Tukey's test for comparing all possible pairs of group means resulted in the following P values (a dash indicates no significant difference):

<table>
<thead>
<tr>
<th>DEPENDENT VARIABLE</th>
<th>RELATIONAL vs RELATIONAL</th>
<th>HIERARCHICAL vs NETWORK</th>
<th>HIERARCHICAL vs NETWORK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Navigational</td>
<td>P &lt; .01</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Specification</td>
<td>P &lt; .05</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

In evaluating the scores on the specification portion of the queries, the cell mean of .78 scored by the Relational group was significantly better than the mean of 2.23 scored by the group using the Hierarchical data model.
In terms of database navigation, the cell mean (2.00) for the experimental group using the Relational Model was significantly better than the mean performance (3.64) of the group using the Hierarchical Model. This finding results in the rejection of Hypothesis la in terms of attribute specification in query writing and in database navigation.

Hypothesis lb states: "Casual user confidence in writing database queries will be the same for each data model used." The results of the ANOVA test, also using Experimental Model 1, indicate no significant main effect on Factor A, data models. This finding does not support the rejection of Hypothesis lb. In summary, the conclusion reached from testing Hypotheses la and lb is that the subjects using the relational model performed significantly better than the subjects using the hierarchical model when writing the specification and navigation portions of the queries, but the level of user confidence is not effected by the data model used.

Hypotheses 2a through 2f are designed to evaluate the effects of the two cognitive style dimensions, Sensing/Intuition and Thinking/Feeling, on casual user's performance and confidence in writing database queries. Hypothesis 2a states: "Casual user performance in writing database queries will be the same for both levels of the Sensing/Intuition cognitive style dimension." The results of ANOVA using the experimental model presented in Figure 4
indicates a significant difference in the number of syntactical errors made when comparing the "S" types to the "N" types. The mean performance of 1.8 computed for the N's is significantly better than the mean of 3.0 recorded by the S types. The results of the ANOVA test for all other aspects of performance failed to identify any significant difference between these two Jungian classifications. The results of evaluating performance using Factor B in Experimental Model 1 are as follows:

<table>
<thead>
<tr>
<th>DEPENDENT VARIABLE</th>
<th>INTUITION vs SENSING</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language Use</td>
<td>P &lt; .05</td>
</tr>
</tbody>
</table>

Hypothesis 2b states: "Casual user confidence in writing database queries will be the same for both levels of the Sensing/Intuition cognitive style dimension." Testing Hypothesis 2b using Experimental Model 1 produced no significant main effect on Factor B, S/N dimension.

Hypothesis 2c states: "Casual user performance in writing database queries will be the same for both levels of the Thinking/Feeling cognitive style dimension." Testing Hypothesis 2c using Experimental Model 2 (Figure 5) produced no significant main effect on Factor B, T/F dimension, for any of the performance aspects.

Hypothesis 2d states: "Casual user confidence in writing database queries will be the same for both levels of
the Thinking/Feeling cognitive style dimension." Using Experimental Model 2 to test this hypothesis indicates a significant difference in user confidence between the T and F types. The mean confidence level of 3.4 for F types was significantly less than the mean of 4.1 reported by the T types (\( \alpha = .01 \)). Thus while the T types did not perform significantly better than the F types, they were more confident.

Hypothesis 2e states: "Casual user performance in writing database queries will be the same for all combinations of levels within the two cognitive style dimensions, Sensing/Intuition and Thinking/Feeling." The use of Experimental Model 3 (Figure 6) in evaluating performance produced no significant main effect on Factor B (Personality Types).

Hypothesis 2f states: "Casual user confidence in writing database queries will be the same for all combinations of levels within the two cognitive style dimensions, Sensing/Intuition and Thinking/Feeling." The use of Experimental Model 3 in evaluating user confidence does indicate a significant main effect on Factor B (Personality Types). This factor contains levels for all combinations of personality types within the two Jungian dimensions being evaluated. Further analysis using Tukey's test indicates the significant difference is between the NT types (cell mean 4.5) and the SF types (cell mean 3.3).
These findings lead to the rejection of Hypothesis 2f.

In summary, the findings resulting from the tests on the second set of hypotheses are: (1) the Intuition types performed significantly better than the Sensing types in terms of the "language use" aspect, (2) the Thinking types were significantly more confident than the Feeling types, and (3) the Intuition/Thinking types were significantly more confident than the Sensing/Feeling types.

Hypotheses 3a through 3f are designed to evaluate the interaction effects of data models and cognitive style classifications on casual user's performance and confidence in writing database queries. These six hypotheses were tested as outlined in Table I. No significant interactions were identified for either user performance or user confidence. Consequently, none of the hypotheses (3a through 3f) could be rejected.
5.1 Summary

This research was motivated by the fact that today's database query facilities are not adequately meeting the needs of the increasing population of casual users. The problem of effectively meeting the needs of this group of users is complicated by: (1) differences in the individuals, (2) variations in the user environments, (3) variations in query facility characteristics, and (4) the wide variety of user needs.

To increase our understanding of the effects of these variables on system usability, which will lead to an eventual solution to this problem, the objectives of this research are twofold. The first objective was to evaluate the effect of each of the three common data models, used for presenting the external view of the database, on the casual user's confidence and ability to write correct database queries. The second objective was to identify cognitive characteristics which influence the individual's confidence and ability to write correct database queries.

In conducting this experiment, sixty-three MBA students with little or no computing background were given the MBTI test. From the resulting cognitive style classifications,
the subjects were randomly selected for one of the three experimental groups. Each group was given an identical description of the contents of a sales-related database and asked to write the same three queries. However, different data models were used to present the external view of the database for each of the three experimental groups. Group 1 received the relational data model, group 2 received the hierarchical data model, and group 3 received the network data model.

All subjects received the same set of rules and examples explaining the use of the English-like command to be written in retrieving information. After writing each of the three queries, the subjects were asked to indicate their confidence in its accuracy. The user's confidence level was reported on a scale of one to five.

The dependent variable 'performance' is represented by the sum of the scores received for the three queries. However, each query was evaluated and received a score based upon five aspects. These scores are as follows: (1) a score for the specification portion of the query, (2) a score for the condition portion of the query, (3) a score for the navigation portion of the query, (4) a score for language use (syntax errors, not considered in the first three scores), and (5) a score for overall performance (this is the sum of the first four scores).
5.2 Discussion of Results

The data collected in this experiment were analyzed using the ANOVA to test for differences in mean user performance and confidence among the three experimental groups. The factorial design uses three experimental models presented and discussed in Section 4.5. Significant main effects on factors involving more than two levels were further analyzed using Tukey's test for comparing all possible pairs of group means. A summary of the statistically significant findings and univariate tests is presented in Table XIII.

<table>
<thead>
<tr>
<th>HYP.</th>
<th>EXP. MODEL</th>
<th>DEPENDENT VARIABLE</th>
<th>TESTING FOR THE EFFECT OF</th>
<th>SIG.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>1</td>
<td>Specification</td>
<td>Relational vs. Hierarchical</td>
<td>*</td>
</tr>
<tr>
<td>1a</td>
<td>1</td>
<td>Navigation</td>
<td>Relational vs. Hierarchical</td>
<td>**</td>
</tr>
<tr>
<td>2a</td>
<td>1</td>
<td>Lang. Use</td>
<td>N Types vs. S Types</td>
<td>*</td>
</tr>
<tr>
<td>2d</td>
<td>2</td>
<td>Confidence</td>
<td>T Types vs. F Types</td>
<td>**</td>
</tr>
<tr>
<td>2f</td>
<td>3</td>
<td>Confidence</td>
<td>NT Types vs. SF Types</td>
<td>*</td>
</tr>
</tbody>
</table>

(*) P < .05  
(**) P < .01
The first set of hypotheses were designed to evaluate the effect of the data models on user's performance and confidence. In the experiment the group using the relational model performed significantly better than the group using the hierarchical model when writing the specification portion and when writing the navigation portion of the query. This finding differs slightly from the findings in the Brosey and Shniederman (1) study. Brosey and Shniederman determined that memorization and in some cases comprehension of the database contents were significantly better using the hierarchical model, compared to the relational model. However, the dependent variables (memorization vs query writing) were quite different in the two experiments, and the Brosey and Shniederman study used a less involved external view of the database. In this study the external view presented with the hierarchical model requires the use of three tree structures, whereas the Brosey and Shniederman external view was a single tree structure.

The second set of hypotheses are designed to evaluate differences in user performance and differences in the user's confidence level among the various cognitive style classifications. The significantly better performance of the Intuition types, compared to the Sensing types, in terms of "language use" was counter intuitive. The literature has portrayed systems designers as typically Sensing types and
users as Intuition types. If this is true, one might expect those having characteristics similar to the designers to perform better. However, an explanation for this difference might be that Sensing types "rely on known facts" and the Intuition types tend to "focus on concepts." In this experiment the subjects were given only one hour to become familiar with the data and learn how to retrieve information. It may be that the Sensing types would have performed better given more time to study the syntax of the retrieval command. Because the retrieval command was English-like, the Intuitive types may have been better able, through the use of reasoning, to grasp the format.

This finding may have other implications, because unlike the typical stereotype, Kaiser and Bostrom (4) found a significantly higher percentage of Sensing types, compared to Intuition types, among the user population. However, the sixty-three MBA students used in this study were split into 71% Sensing types and 29% Intuition types. This could be a good indication that there are significantly more Sensing types than Intuition types in the population as a whole, or at least within the population of business representatives. On the other hand, Kiser and Bostrom found that systems designers were fairly balanced on this dimension (Sensing/Intuition). This could indicate that the fewer number of Intuition types have an attraction to the design occupations.
Ghani (2) determined that the Feeling types preferred and performed better with graphic presentations, while the Thinking types preferred and performed better with tabular presentations. In this study the Thinking types were significantly more confident than the Feeling types. This was also supported by the fact that the NT types were more confident than the SF types. These two personality types (NT vs SF) are extreme opposites (see Figure 2 - Model of Jungian Personality Types). To support Ghani's findings, however, one might expect the Thinking types to perform better and be more confident with the relational model, and the Feeling types to perform better and be more confident with the hierarchical and network models. This was not the case, no significant differences were detected within the data model groups.

The third set of hypotheses are designed to evaluate any interaction effects among the levels of the two independent variables. The absence of any interaction effects between data models (Factor A) and cognitive style classifications (Factor B) support the findings in Hoffer's study. Hoffer (3) used the Klob Learning Style Inventory in a data model related experiment, and concluded that "cognitive style need not be considered in selecting a data model." The findings from this study support that conclusion.
5.3 Implications of the Findings

The purpose of this study is to extend the basic knowledge relating to database query facilities ease-of-use for casual users. The research questions introduced in Section 1.3 have been answered.

The findings from this study will benefit both developers of DBMS and database systems designers within the user community. Developers of DBMS should recognize the benefits of providing a system that will support a database that can be presented to the casual users in tabular form. They should also provide training that will aid systems designers in developing appropriate documentation of the database for casual users.

Systems designers must recognize that to achieve casual user acceptance, two ingredients are required by the system. First the casual user must be able to use the system effectively. Secondly, the casual user must be confident in his ability to use the system and the reliability of the information being retrieved.

The results of this study indicate that using the relational model to present the external view will increase casual user performance in writing database queries. Other studies have determined that both performance and confidence was better when using the hierarchical data model. However, these studies used non-query writing tasks. This could be an indication that some of the dependent variables used in
previous studies to measure ease-of-use do not reflect the
users' ability to write database queries.

The study also indicates that cognitive style need not
be considered in the selection of the data model used in
presenting the external view. However, it is indicated that
Feeling personality types need more support than Thinkink
types to increase their confidence level.

5.4 Further Research

Because of the variations between the findings of this
study and the Brosey and Shniederman study, further research
is needed to evaluate database data model effectiveness for
casual users. Possibly, a replication of this study in a
business setting would be appropriate. This would permit
the use of subjects already familiar with the database
contents, eliminating one of the limitations of this study.
The business setting could also provide a more realistically
involved external view of the DB. Research is also needed
to evaluate the effectiveness of tasks which are being used
to measure ease-of-use. This study has raised some doubt as
to whether the dependent variables used in other studies are
appropriate for measuring database success.

Query development in this study was totally off-line.
A replication of this study using an interactive database
system would more closely approximate the casual user's
actual environment.

Procedural vs. nonprocedural languages is one of the
major issues affecting query facilities ease-of-use. Welty (5), conducted a study comparing SQL to TABLET. He determines that SQL is more procedural than TABLE, through the use of a "procedurality metric". The findings of the study indicate that the procedural language is best when solving more difficult problem, and for the more advanced users. However, he suggests that the nonprocedural language (SQL) may be best for less experienced users. Based upon Welty's "procedurality metric", the query command used in this study was less procedural than SQL. A future study might investigate interaction effects between data models and different levels of procedurality.

Additional research is needed to further investigate the effects and characteristics of the Sensing/Intuition personality dimension as measured by the MBTI. Is there really a difference in performance between these two Jungian classifications? What are the characteristics of a typical user population? What type of dialogue would be better suited for Intuition Types? How can systems be designed to increase the confidence level of Feeling Types?

As Hoffer (3) suggests, other cognitive aspects need to be investigated. For example, what influence does cognitive complexity have on casual user's performance and data model effectiveness?
CHAPTER BIBLIOGRAPHY


APPENDIX A

EXERCISE BOOKLET A (RELATIONAL DATA MODEL)
INTRODUCTION

You have been asked to participate in an information retrieval exercise designed to evaluate the performance of managers in a business environment. The exercise is presented in two booklets. The first booklet contains (1) a very brief introduction to the business environment, (2) a description of the data available to management, (3) a description of the rules for retrieving information, and (4) two practice problems. One hour will be allowed to study Booklet One and ask questions. All questions will be resolved during the one-hour study period ONLY.

In the second booklet you will be asked to write instructions to retrieve specific information. When writing the instructions in Booklet Two you will be allowed to refer back to this booklet which contains the procedure for retrieving information and examples. Therefore, while studying this material, emphasis should be on understanding the data and how to retrieve information rather than on memorizing any of the material.

PART I - INTRODUCTION TO SPECIAL ELECTRIC COMPANY

The company assembles and markets select electrical components at a reasonable profit. The components to be manufactured are selected because of their wide variety of uses in industry. Because of this wide usage, the company can produce large quantities during a given production run and maintain a large finished-goods inventory to provide prompt customer deliveries.
The managerial decisions in this exercise involve four major entities. Figure 1 illustrates these entities and the relationships among them.

**FIGURE 1 - MANAGEMENT VIEW OF SALES RELATED DATA**

The entities defined in this view are connected with lines representing the relationships among the entities. The relationship between "customers" and "customer orders" (number 1) is one-to-many. This implies that a customer may place any number of orders, but that a given order will pertain to only one customer. The relationship between "customer orders" and "finished-goods inventory" (number 2), like that between "parts inventory" and "finished-goods inventory" (number 3), is many-to-many. This implies that a customer order may contain any number of components from finished-goods inventory (different
line items) and that a given component in inventory may appear on any number of different orders. Relationship number 3 implies that a given component in finished-goods inventory is assembled from any number of different parts and that a given part may be used in the construction of any number of different electrical components.

PART II - DATA DESCRIPTION

This part provides a detailed description of the data used by sales management for decision making. The data pertaining to the four entities in the management view (Figure 1) are defined in six tables. Figure 2 identifies the table names and the column names within each table. Figure 3 provides an example of the kind of data found in each table. For example, a row in the customer master table (CUST-MST) will contain information about one customer, while a row in the finished-goods inventory table (FG-INV) will contain information about the inventory on one electrical component.

During the exercise, you will need to refer to Figures 2 and 3 frequently. To make this referral easier, these figures are provided as separate handouts. Please refer to these handouts when studying the general descriptions of the table and column names which follow:

1 - The Customer Master Table contains columns of customer number, customer name, salesman code, current year sales, and prior year sales.

2 - The Customer Orders Table contains columns of order number, customer number, date the order was placed, and the date the
FIGURE 2 - EXTERNAL VIEW (RELATIONAL MODEL)

<table>
<thead>
<tr>
<th>Table Name</th>
<th>Column Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUST-MST</td>
<td>(CUST-NO, CUST-NAME, SM-CODE, C-YR-SALES, P-YR-SALES)</td>
</tr>
<tr>
<td>CUST-ORD</td>
<td>(ORD-NO, CUST-NO, ORO-DATE, SHIP-DATE)</td>
</tr>
<tr>
<td>ORD-LI</td>
<td>(ORD-NO, LINE-NO, COMP-NO, ORD-UNIT-PR, ORD-QTY)</td>
</tr>
<tr>
<td>FG-INV</td>
<td>(COMP-NO, COMP-DESC, FG-REORD-PT, FG-UNIT-PR, FG-QTY)</td>
</tr>
<tr>
<td>PART-INV</td>
<td>(PART-NO, PART-DESC, PART-REORD-PT, PART-UNIT-PR, PART-QTY)</td>
</tr>
<tr>
<td>MAT-EXP</td>
<td>(COMP-NO, PART-NO, EXP-QTY)</td>
</tr>
</tbody>
</table>
**Figure 3 -- Sample Relations**

### Cust-Ord

<table>
<thead>
<tr>
<th>Cust-No</th>
<th>Cust-Name</th>
<th>SM-Code</th>
<th>C-Yr-Sales</th>
<th>P-Yr-Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>524</td>
<td>General Electric Corp.</td>
<td>03</td>
<td>1,291,731</td>
<td>945,772</td>
</tr>
<tr>
<td>526</td>
<td>Bowling A/C Co.</td>
<td>03</td>
<td>652,491</td>
<td>483,621</td>
</tr>
</tbody>
</table>

### Part-Inv

<table>
<thead>
<tr>
<th>Part-No</th>
<th>Part-Desc</th>
<th>Part-Reord-Pr</th>
<th>Part-Unit-Pr</th>
<th>Part-Qty</th>
</tr>
</thead>
<tbody>
<tr>
<td>605</td>
<td>DD Armatures</td>
<td>30</td>
<td>60.30</td>
<td>41</td>
</tr>
<tr>
<td>614</td>
<td>Copper Wire No</td>
<td>1000</td>
<td>.75</td>
<td>1209</td>
</tr>
<tr>
<td>634</td>
<td>Spring III</td>
<td>300</td>
<td>.15</td>
<td>409</td>
</tr>
<tr>
<td>652</td>
<td>Coil 12</td>
<td>80</td>
<td>9.90</td>
<td>87</td>
</tr>
<tr>
<td>644</td>
<td>Bearing Pin 31</td>
<td>90</td>
<td>1.45</td>
<td>10</td>
</tr>
<tr>
<td>652</td>
<td>Back Stop III</td>
<td>1200</td>
<td>12.20</td>
<td>1332</td>
</tr>
</tbody>
</table>

### Mat-Exp

<table>
<thead>
<tr>
<th>Comp-No</th>
<th>Part-No</th>
<th>Exp-Qty</th>
</tr>
</thead>
<tbody>
<tr>
<td>851</td>
<td>638</td>
<td>3</td>
</tr>
<tr>
<td>851</td>
<td>644</td>
<td>5</td>
</tr>
<tr>
<td>852</td>
<td>644</td>
<td>7</td>
</tr>
<tr>
<td>852</td>
<td>614</td>
<td>1</td>
</tr>
<tr>
<td>852</td>
<td>652</td>
<td>2</td>
</tr>
<tr>
<td>850</td>
<td>605</td>
<td>1</td>
</tr>
<tr>
<td>850</td>
<td>614</td>
<td>4</td>
</tr>
<tr>
<td>850</td>
<td>634</td>
<td>2</td>
</tr>
<tr>
<td>850</td>
<td>652</td>
<td>1</td>
</tr>
</tbody>
</table>

### 3rd-Li

<table>
<thead>
<tr>
<th>Cust-No</th>
<th>Cust-Name</th>
<th>SM-Code</th>
<th>C-Yr-Sales</th>
<th>P-Yr-Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>197</td>
<td>524</td>
<td>03</td>
<td>1,291,731</td>
<td>945,772</td>
</tr>
<tr>
<td>197</td>
<td>526</td>
<td>03</td>
<td>652,491</td>
<td>483,621</td>
</tr>
</tbody>
</table>

### 3rd-Li

<table>
<thead>
<tr>
<th>Cust-No</th>
<th>Cust-Name</th>
<th>SM-Code</th>
<th>C-Yr-Sales</th>
<th>P-Yr-Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>197</td>
<td>524</td>
<td>03</td>
<td>1,291,731</td>
<td>945,772</td>
</tr>
<tr>
<td>197</td>
<td>526</td>
<td>03</td>
<td>652,491</td>
<td>483,621</td>
</tr>
</tbody>
</table>

### FG-Inv

<table>
<thead>
<tr>
<th>Comp-No</th>
<th>Comp-Desc</th>
<th>FG-Reord-Pr</th>
<th>FG-Unit-Pr</th>
<th>FG-Qty</th>
</tr>
</thead>
<tbody>
<tr>
<td>851</td>
<td>Diliight Selector Switch (STM)</td>
<td>400</td>
<td>35.10</td>
<td>317</td>
</tr>
<tr>
<td>852</td>
<td>(WL)</td>
<td>950</td>
<td>95.10</td>
<td>402</td>
</tr>
<tr>
<td>853</td>
<td>Time Switch (SPST) - 120V</td>
<td>300</td>
<td>128.07</td>
<td>321</td>
</tr>
</tbody>
</table>
components are to be shipped. This table contains header information for the customer order, with the individual line items being stored in the next table.

3 - The Order Line Item Table contains columns of order number, line number, component number, sales price per component, and the number of components ordered.

4 - The Finished-Goods Inventory Table contains columns of component number, component description, the inventory level at which additional production should be ordered (finished-goods reorder point), sales price per component, and quantity on hand.

5 - The Parts Inventory Table contains columns of part number, part description, part reorder point, purchase price per part, and quantity on hand.

6 - The Materials Explosion Table contains columns of component number, part number, and the number of these parts required to build one component (explosion quantity). This quantity represents the quantity of this specific part required to assemble a given electrical component.

The exercise in Booklet Two will require a thorough understanding of the different tables, their contents, and interrelationships. The interrelationship between tables is established through the use of corresponding columns. For example, looking at the CUST-MST and CUST-ORD tables in Figure 3, we can see that order number 197 belongs to General Electric Corporation because of the corresponding customer number, 524.

Due to space limitations, Figure 3 shows only a small portion of the actual data that exists for each table within the system. However, with these examples, Figure 3 can be used to gain an understanding of the existing relationships between tables. The next two examples illustrate the interrelationships among the tables in the system. Please trace these examples carefully, using Figure 3.
Example 1

1 - Looking at the CUST-MST and CUST-ORD tables, we can see that Bowing A/C Co., customer master number 516, has two outstanding orders, order numbers 116 and 161.

2 - In the order line item table we can see that order number 161 has two line items, numbers 01 and 02.

3 - Line item 01 is for eighty-three units of component number 851. Looking at FG-INV, we can see that component number 851 is Oiltight Selector Switches (STD) and the inventory balance is 317.

4 - Looking at the MAT-EXP table, we can see that this component (851) is assembled using three of part number 638 and five of part number 644. From the PART-INV table, we can see that part number 644 is described as "BEARING PIN 31," with a quantity of 30 on-hand.

Example 2

1 - From materials explosion we can see that part number 614 is used in the construction of both component numbers 852 and 850. Construction of component number 850 requires four of part number 614.

2 - Looking at ORD-LI, we can see that there are three outstanding orders for component 850. These are line 02 of order 161, line 02 of order 116, and line 03 of order 197. From the CUST-ORD table we can see that order number 197 is for customer number 524.

3 - Looking at CUST-MST, we can see that 524 is General Electric Corp.

The purpose of these examples is to illustrate how the six tables are interrelated through the use of corresponding columns.

PART III - RULES FOR WRITING INFORMATION RETRIEVAL REQUESTS

This part is divided into two sections. The first section contains general requirements which must be met when writing requests and definitions of key terms used in the information retrieval command. The second section describes the LIST
command that is used in a request, along with examples of how this command can be used. Use Figure 3 as an aid to understanding the examples.

GENERAL REQUIREMENTS AND DEFINITIONS

1 - Precise spelling must be used for all column names and table names (see Figure 2 or Figure 3 for these spellings).

2 - A reference to a specific column must contain both the column name and the table name.

Example 1 - CUST-NAME in CUST-MST

Example 2 - CUST-NO, CUST-NAME, and SM-CODE in CUST-MST

The second example illustrates that the table name need not be repeated when referencing multiple columns within the same table.

3 - A condition may be defined in the LIST command by using the "when" clause. The format of a condition is as follows:

\[
\text{when } \begin{cases} \{ < \} & \{ \text{reference 2} \} \\ \{ = \} & \{ \text{or a value} \} \end{cases}.
\]

Compound conditions may be defined using the "and" and "or" conjunctions.

Example 1 - when C-YR-SALES in CUST-MST < 500,000

Example 2 - when P-YR-SALES in CUST-MST > C-YR-SALES in CUST-MST and C-YR-SALES in CUST-MST > 1,000,000

DESCRIPTION OF LIST COMMAND

As an aid to understanding the purpose of this command, try to visualize the results of using the command while studying the following examples.

LIST - This command provides the capability of printing lines using specific columns from one or more tables (see example 1 below). The line or lines to be printed may be based upon an optional condition through the use of the
"when" clause (see example 2 below). In addition, you must use one or more "THRU" clauses at the end of the list command when relating to two tables that do not have corresponding columns (see example 3 below).

Example 1 - LIST COMP-NO and COMP-DESC in FG-INV.

This example will print the component number and description for all rows in the finished-goods inventory table.

RESULTS:       COMP-NO       COMP-DESC
               851       OILTIGHT SELECTOR SWITCH (STD)
               852       OILTIGHT SELECTOR SWITCH (WL)
               850       TIME SWITCH (SPST) - 120V

Example 2 - LIST COMP-DESC in FG-INV and EXP-QTY in MAT-EXP when PART-NO in MAT-EXP = 614.

In this example the system will search MAT-EXP until the condition in the "when" clause is met. It will then print the EXP-QTY for this row and the corresponding COMP-DESC from FG-INV. column in the two tables, COMP-NO.

Write the results of this request in the space which follows:

Check this answer with answer 1 on the last page.

Example 3 - LIST CUST-NAME in CUST-MST and ORD-QTY in ORD-LI when ORD-QTY in ORD-LI > 100 THRU CUST-ORD.
In this example the system will search ORD-LI until the condition in the "when" clause is met. It will then locate the corresponding row in the CUST-ORD table (because of the THRU clause). From here it will locate and print the corresponding customer name in the customer master table, along with the ORD-QTY which met the original condition. It is important to note that the direction of the linkage is from the table specified in the "when" clause to the information being printed!

Write the results of this request in the space which follows:

Check this answer with answer 2 on the last page.

Following are four examples of complete information retrieval requests which could be used with this system. Pay particular attention to the translation of the English statement of the problem into the request. Similar translations will be required in the writing exercise contained in Booklet Two.

--------------------------- EXAMPLE 1 ---------------------------

English Statement of Problem: Which finished-goods inventory items are priced less than 100 dollars?

Information Retrieval Request: List COMP-NO and COMP-DESC in FG-INV when FG-UNIT-PR in FG-INV < 100.00.
The results from this request are:

<table>
<thead>
<tr>
<th>COMP-NO</th>
<th>COMP-DESC</th>
</tr>
</thead>
<tbody>
<tr>
<td>851</td>
<td>OILTIGHT SELECTOR SWITCH (STD)</td>
</tr>
<tr>
<td>852</td>
<td>OILTIGHT SELECTOR SWITCH (WL)</td>
</tr>
</tbody>
</table>

This request requires the use of only one table "FG-INV." All requests must be explicit about the table(s) and column(s) being referenced. Refer to Figure 3 to see how this request for information was satisfied.

------------------------- EXAMPLE 2 -------------------------

**English Statement of Problem:** What customers are scheduled for shipment on 10-12-83?

**Information Retrieval Request:** List CUST-NO and CUST-NAME in CUST-MST when SHIP-DATE in CUST-ORD = 10-12-83.

**Results:**

<table>
<thead>
<tr>
<th>CUST-NO</th>
<th>CUST-NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>516</td>
<td>BOWING A/C CO.</td>
</tr>
</tbody>
</table>

This request requires the use of two tables which have a corresponding column, CUST-NO. No THRU clause is required.

------------------------- EXAMPLE 3 -------------------------

**English Statement of Problem:** Which finished-goods components are assembled using parts that are below the part reorder point?

**Information Retrieval Request:** LIST COMP-NO and COMP-DESC in FG-INV and PART-NO in PART-INV when PART-QTY in PART-INV < PART-REORD-PT in PART-INV, THRU MAT-EXP.
This request involves the use of two tables, FG-INV and PART-INV. However, these two tables do not contain corresponding columns. Because of this, the THRU clause was used to define the linkage from PART-INV through MAT-EXP to the FG-INV table. Note that the THRU clause must be used at the end of the request when referencing a second table that does not contain a corresponding column.

EXAMPLE 4

English Statement of Problem: Which parts are used in the construction of components ordered for Bowing A/C Co.?

Information Retrieval Request: LIST PART-NO and PART-DESC in PART-INV when CUST-NAME in CUST-MST = Bowing A/C Co., THRU CUST-ORD, THRU ORD-LI, THRU MAT-EXP.

Results:

<table>
<thead>
<tr>
<th>PART-NO</th>
<th>PART-DESC</th>
</tr>
</thead>
<tbody>
<tr>
<td>606</td>
<td>DD ARMATURES</td>
</tr>
<tr>
<td>614</td>
<td>COPPER WIRE 40</td>
</tr>
<tr>
<td>634</td>
<td>SPRING 111</td>
</tr>
<tr>
<td>652</td>
<td>BACK STOP 111</td>
</tr>
<tr>
<td>638</td>
<td>COIL 12</td>
</tr>
<tr>
<td>644</td>
<td>BEARING PIN 31</td>
</tr>
</tbody>
</table>

This request involves the use of two tables, PART-INV and CUST-MST. Also, as in example 3, these two tables do not contain corresponding columns. Because of this, the THRU clause was used
to define the linkage from CUST-MST to PART-INV. However, in this example, you must define the navigation from CUST-MST through CUST-ORD (corresponding column is CUST-NO), and from CUST-ORD through ORD-LI (corresponding column is COMP-NO), and then from ORD-LI through MAT-EXP (corresponding column is COMP-NO) to get to PART-INV.
PART IV - PRACTICE PROBLEMS

In the two practice problems which follow, you are to write the request and then show the results of the request.

**English Statement of Practice Problem 1** - What is the inventory balance for component number 852? Show the component description and the inventory balance.

**Information Retrieval Request 1** - (Check with answer 3)

Results:

**English Statement of Practice Problem 2** - What parts are used in the construction of components that have a sales price of less than fifty dollars? Show only the part description for these parts.

**Information Retrieval Request 2** - (Check with answer 4)

Results:
## ANSWERS TO PRACTICE PROBLEMS

### ANSWER 1 -

<table>
<thead>
<tr>
<th>COMP-DESC</th>
<th>EXP-QTY</th>
</tr>
</thead>
<tbody>
<tr>
<td>OILTIGHT SELECTOR SWITCH (WL)</td>
<td>1</td>
</tr>
<tr>
<td>TIME SWITCH (SPST) - 120V</td>
<td>4</td>
</tr>
</tbody>
</table>

### ANSWER 2 -

<table>
<thead>
<tr>
<th>CUST-NAME</th>
<th>ORD-QTY</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOWING A/C CO.</td>
<td>112</td>
</tr>
</tbody>
</table>

### ANSWER 3 -

**LIST COMP-DESC and FG-QTY in FG-INV when COMP-NO in FG-INV = 852.**

**RESULTS:**

<table>
<thead>
<tr>
<th>COMP-DESC</th>
<th>FG-QTY</th>
</tr>
</thead>
<tbody>
<tr>
<td>OILTIGHT SELECTOR SWITCH (WL)</td>
<td>492</td>
</tr>
</tbody>
</table>

### ANSWER 4 -

**LIST PART-DESC in PART-INV when FG-UNIT-PR in FG-INV < 50.00, THRU MAT-EXP.**

**RESULTS:**

<table>
<thead>
<tr>
<th>PART-DESC</th>
</tr>
</thead>
<tbody>
<tr>
<td>COIL 12</td>
</tr>
<tr>
<td>BEARING PIN 31</td>
</tr>
</tbody>
</table>
APPENDIX B

EXERCISE BOOKLET B (HIERARCHICAL DATA MODEL)
INTRODUCTION

You have been asked to participate in an information retrieval exercise designed to evaluate the performance of managers in a business environment. The exercise is presented in two booklets. The first booklet contains (1) a very brief introduction to the business environment, (2) a description of the data available to management, (3) a description of the rules for retrieving information, and (4) two practice problems. One hour will be allowed to study Booklet One and ask questions. All questions will be resolved during the one-hour study period ONLY.

In the second booklet you will be asked to write instructions to retrieve specific information. When writing the instructions in Booklet Two you will be allowed to refer back to this booklet which contains the procedure for retrieving information and examples. Therefore, while studying this material, emphasis should be on understanding the data and how to retrieve information rather than on memorizing any of the material.

PART I - INTRODUCTION TO SPECIAL ELECTRIC COMPANY

The company assembles and markets select electrical components at a reasonable profit. The components to be manufactured are selected because of their wide variety of uses in industry. Because of this wide usage, the company can produce large quantities during a given production run and maintain a large finished-goods inventory to provide prompt customer deliveries.
The managerial decisions in this exercise involve four major entities. Figure 1 illustrates these entities and the relationships among them.

**FIGURE 1 - MANAGEMENT VIEW OF SALES RELATED DATA**

The entities defined in this view are connected with lines representing the relationships among the entities. The relationship between "customers" and "customer orders" (number 1) is one-to-many. This implies that a customer may place any number of orders, but that a given order will pertain to only one customer. The relationship between "customer orders" and "finished-goods inventory" (number 2), like that between "parts inventory" and "finished-goods inventory" (number 3), is many-to-many. This implies that a customer order may contain any number of components from finished-goods inventory (different
line items) and that a given component in inventory may appear on any number of different orders. Relationship number 3 implies that a given component in finished-goods inventory is assembled from any number of different parts and that a given part may be used in the construction of any number of different electrical components.

PART II - DATA DESCRIPTION

This part provides a detailed description of the data used by sales management for decision making. The data pertaining to the four entities in the management view (Figure 1) are defined in eight record types. These eight record types are grouped into three hierarchies (or trees). Figure 2 identifies the record type names and the field names within each record type. Figure 3 provides an example of the kind of data found in each record type. For example a customer master record (CUST-MST) will contain information about one customer, while a record in finished-goods inventory (FG-INV) will contain information about the inventory on one electrical component.

During the exercise, you will need to refer to Figures 2 and 3 frequently. To make this referral easier these figures are provided as separate handouts. Please refer to these handouts when studying the general descriptions of the record and field names which follow:

----- TREE 1 -----

1 - The Customer Master record type contains fields of customer number, customer name, salesman code, current year sales, and prior year sales.

2 - The Customer Orders record type contains fields of order number, date the order was placed, and the date the
FIGURE 2 -- EXTERNAL VIEW (HIERARCHICAL MODEL)

(TREE 1)

CUST-MST

CUST-NO CUST-NAME SM-CODE C-YR-SALES P-YR-SALES

CUST-ORD

ORD-NO ORD-DATE SHIP-DATE

ORD-LI

LINE-NO COMP-NO ORD-UNIT-PR ORD-QTY

(TREE 2)

FG-INV

COMP-NO COMP-DESC FG-REORD-PT FG-UNIT-PR FG-QTY

FG-LI

CUST-NO ORD-NO LINE-NO

MAT-EXP

PART-NO EXP-QTY

(TREE 3)

PART-INV

PART-NO PART-DESC PART-REORD-PT PART-UNIT-PR PART-QTY

PART-REQ

COMP-NO
FIGURE 3 -- SAMPLE HIERARCHIES

(FIGURE 1)

CUST-INV
524 GENERAL ELECTRIC CORP. 01 1,243,741 946,772
616 ELECTRIC MACH CO. 09 562,495 482,621

CUST-ORD
197 19-08-32 10-14-32
115 10-10-93 10-12-32
161 10-12-87 10-17-82

ORD-L1
01 851 28.97 125
02 852 92.93 103
02 855 137.02 132

(FIGURE 2)

FG-INV
851 OILTIGHT SELECTOR SWITCH (STD) 802 35.10 317
852 OILTIGHT SELECTOR SWITCH (HL) 802 95.10 402
850 TIME SWITCH (SPST) 723V 330 125.0 321

FG-L1
524 197 02
516 116 02
516 117 02

MAT-EXP
524 197 02
552 2
514 2

(FIGURE 3)

PART-INV
606 DD ARMATURES 30 60.30 41

614 COPPER WIRE 40 103 15 1229
614 SPRING 111 300 15 400
638 COIL 12 82 9.93 87
544 BEARING PIV 31 90 3.45 30
652 BACK STOP 111 128 12.20 1332

PART-REQ
850
852
851 852

850
components are to be shipped. This record type contains header information for the customer order, with the individual line items being stored in the next record type.

3 - The Order Line Item record type contains fields of line number, component number, sales price per component, and the number of components ordered.

------ TREE 2 ------

4 - The Finished-Goods Inventory record type contains fields of component number, component description, the inventory level at which additional production should be ordered (finished-goods reorder point), sales price per component, and quantity on hand.

5 - The Finished-Goods Line Item record type contains customer number, order number, and line number. This record type is used to establish a linkage back to Tree 1 through the use of corresponding fields. For example, with the customer number in the FG-LI record type, we are able to get to the customer records in CUST-MST that have this component on order.

6 - The Materials Explosion record type contains part number and explosion quantity. This quantity represents the quantity of this specific part required to assemble a given electrical component.

------ TREE 3 ------

7 - The Parts Inventory record type contains fields of part number, part description, part reorder point, purchase price per part, and quantity on hand.

8 - The Parts Requirements record type contains a component number field only. This allows access to the finished-goods inventory records in which a particular part is used.

The exercise in Booklet Two will require you to have a thorough understanding of the different record types, their contents, and interrelationships. The interrelationship between records at adjacent levels within a hierarchy is defined with connecting lines. The interrelationship between records at non-adjacent levels within a hierarchy is defined with two connecting lines and an intervening record type. For example, the interrelationship between CUST-MST and ORD-LI is through the
The interrelationships between hierarchies is established through the use of corresponding fields. For example, because the component number is in the ORD-LI records, we can get to the specific inventory record in Tree 2.

Due to space limitations, Figure 3 shows only a small portion of the actual data that exists for each record type. However, with these examples, Figure 3 can be used to gain an understanding of the existing relationships. The next two examples illustrate the interrelationships among the record types and hierarchies in the system. Please trace these examples carefully, using Figure 3.

**Example 1**

1 - Looking at the CUST-MST and CUST-ORD record types, we can see that Bowin A/C Co., customer master record number 516, has two orders, order numbers 116 and 161.

2 - Following order number 161 to the ORD-LI records, we see that this order has two line items, numbers 01 and 02.

3 - Line item 01 is for eighty-three units of component number 851. Looking at FG-INV, we can see that component number 851 is Oiltight Selector Switches (STD) and the inventory balance is 317.

4 - Following component 851 to the MAT-EXP record types we can see that this component is assembled using three of part number 638 and five of part number 644. From the PART-INV record type we can see that part number 644 is described as "BEARING PIN 31," with a quantity of 30 on-hand.

**Example 2**

1 - From TREE 3 we can see that part number 614, COPPER WIRE 40, is used in the construction of both component numbers 850 and 852. Looking at FG-INV and MAT-EXP we see the construction of component 850 requires four of part 614.

2 - Looking at FG-LI in TREE 2, we can see that there are three outstanding orders for component 850. These are line 02 of
order 161, line 02 of order 116, and line 03 of order 197. From the CUST-ORD record type we see order number 197 is for customer number 524, General Electric Corp.

3 - Looking down this hierarchy, we can see that line number 03 of order number 197 is for 100 units of component 850.

The purpose of these examples is to illustrate how records are related within a hierarchy and how the three hierarchies are interrelated through the use of corresponding fields.

PART III - RULES FOR WRITING INFORMATION RETRIEVAL REQUESTS

This part is divided into two sections. The first section contains general requirements which must be met when writing requests and definitions of key terms used in the information retrieval command. The second section describes the command that may be used in a request, along with examples of how this command can be used. Use Figure 3 as an aid to understanding the examples.

GENERAL REQUIREMENTS AND DEFINITIONS

1 - Precise spelling must be used for all field names and record names (see Figure 2 for these spellings).

2 - A reference to a specific field must contain both the field name and the record type name.

Example 1 - CUST-NAME in CUST-MST

Example 2 - CUST-NO, CUST-NAME, and SM-CODE in CUST-MST

The second example illustrates that the record type name need not be repeated when referencing multiple fields within the same record type.

3 - A condition may be defined in the LIST command by using the "when" clause. The format of a condition is as follows:

\[
\text{when reference 1} \begin{cases}
< \\
>
\\
= \\
\text{or a value}
\end{cases}
\]

Compound conditions may be defined using the "and" and "or"
conjunctions.

Example 1 - when C-YR-SALES in CUST-MST < 500,000

Example 2 - when P-YR-SALES in CUST-MST > C-YR-SALES in CUST-MST and C-YR-SALES in CUST-MST > 1,000,000

DESCRIPTION OF LIST COMMAND

As an aid to understanding the purpose of this command, try to visualize the results of using the command while studying the following examples.

LIST - This command provides the capability of printing lines using specific fields from one or more record types (see example 1, below). The line or lines to be printed may be based upon an optional condition through the use of the "when" clause (see example 2 below). In addition, you must use one or more "THRU" clauses at the end of the list command when (1) referring to two record types within a hierarchy that are not connected directly, or (2) referring to two record types in two different hierarchies that do not contain corresponding fields (see example 3 below).

Example 1 - LIST COMP-NO and COMP-DESC in FG-INV.

This example will print the component number and description for all finished-goods inventory records.

RESULTS:

<table>
<thead>
<tr>
<th>COMP-NO</th>
<th>COMP-DESC</th>
</tr>
</thead>
<tbody>
<tr>
<td>851</td>
<td>OILTIGHT SELECTOR SWITCH (STD)</td>
</tr>
<tr>
<td>852</td>
<td>OILTIGHT SELECTOR SWITCH (WL)</td>
</tr>
<tr>
<td>850</td>
<td>TIME SWITCH (SPST) - 120V</td>
</tr>
</tbody>
</table>
Example 2 - LIST COMP-DESC in FG-INV and EXP-QTY in MAT-EXP when PART-NO in MAT-EXP = 614.

In this example the system will search MAT-EXP until the condition in the "when" clause is met. It will then print the EXP-QTY for this record and the corresponding COMP-DESC from FG-INV.

Write the results of this request in the space which follows:

Check this answer with answer 1 on the last page.

Example 3 - LIST CUST-NAME in CUST-MST and ORD-QTY in ORD-LI when ORD-QTY in ORD-LI > 100 THRU CUST-ORD.

In this example the system will search ORD-LI until the condition in the "when" clause is met. It will then link to the record in CUST-ORD (because of the THRU clause). From here it will link to the record in CUST-MST and print customer name, along with the ORD-QTY which met the original condition. It is important to note that the direction of the linkage is from the record type specified in the "when" clause to the information being printed!

Write the results of this request in the space which follows:
Check this answer with answer 2 on the last page.

Following are four examples of complete information retrieval requests which could be used with this system. Pay particular attention to the translation of the English statement of the problem into the request. Similar translations will be required in the writing exercise contained in Booklet Two.

--------------------- EXAMPLE 1 ---------------------

**English Statement of Problem:** Which finished-goods inventory items are priced less than 100 dollars?

**Information Retrieval Request:** List COMP-NO and COMP-DESC in FG-INV when FG-UNIT-PR in FG-INV < 100.00.

The results from this request are:

<table>
<thead>
<tr>
<th>COMP-NO</th>
<th>COMP-DESC</th>
</tr>
</thead>
<tbody>
<tr>
<td>851</td>
<td>OILTIGHT SELECTOR SWITCH (STD)</td>
</tr>
<tr>
<td>852</td>
<td>OILTIGHT SELECTOR SWITCH (WL)</td>
</tr>
</tbody>
</table>

This request requires the use of only one record type, "FG-INV." All requests must be explicit about the record type(s) and field(s) being referenced.

--------------------- EXAMPLE 2 ---------------------

**English Statement of Problem:** What customers are scheduled for shipment on 10-12-83?

**Information Retrieval Request:** List CUST-NO and CUST-NAME in CUST-MST when SHIP-DATE in CUST-ORD = 10-12-83.
Results:

<table>
<thead>
<tr>
<th>CUST-NO</th>
<th>CUST-NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>516</td>
<td>BOWING A/C CO.</td>
</tr>
</tbody>
</table>

This request requires the use of two record types at adjacent levels within a hierarchy. No THRU clause is required.

---------------------- EXAMPLE 3 ----------------------

English Statement of Problem: Which finished-goods components are assembled using parts that are below the part reorder point?

Information Retrieval Request: LIST COMP-NO and COMP-DESC in FG-INV and PART-NO in PART-INV when PART-QTY in PART-INV < PART-REORD-PT in PART-INV, THRU PART-REQ.

Results:

<table>
<thead>
<tr>
<th>COMP-NO</th>
<th>COMP-DESC</th>
<th>PART-NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>851</td>
<td>OILTIGHT SELECTOR SWITCH (STD)</td>
<td>644</td>
</tr>
<tr>
<td>852</td>
<td>OILTIGHT SELECTOR SWITCH (WL)</td>
<td>644</td>
</tr>
</tbody>
</table>

This request involves the use of two record types, FG-INV and PART-INV. However, these two record types are located in two different hierarchies and do not contain corresponding fields. Because of this, the THRU clause is used to define the linkage from PART-INV through PART-REQ to the FG-INV record type. Note that the THRU clause must be used at the end of the request when referencing a second record type that is located in a nonadjacent level within the same hierarchy or is in another hierarchy and does not contain a corresponding field.
EXAMPLE 4

English Statement of Problem: Which parts are used in the construction of components ordered for Bowing A/C Co.?

Information Retrieval Request: LIST PART-NO and PART-DESC in PART-INV when CUST-NAME in CUST-MST = Bowing A/C Co. THRU CUST-ORD, THRU ORD-LI, THRU PART-REQ.

Results:

<table>
<thead>
<tr>
<th>PART-NO</th>
<th>PART-DESC</th>
</tr>
</thead>
<tbody>
<tr>
<td>606</td>
<td>DD ARMATURES</td>
</tr>
<tr>
<td>614</td>
<td>COPPER WIRE 40</td>
</tr>
<tr>
<td>634</td>
<td>SPRING 111</td>
</tr>
<tr>
<td>652</td>
<td>BACK STOP 111</td>
</tr>
<tr>
<td>638</td>
<td>COIL 12</td>
</tr>
<tr>
<td>644</td>
<td>BEARING PIN 31</td>
</tr>
</tbody>
</table>

This request involves the use of two record types, PART-INV and CUST-MST. Also, as in example 3, these two record types do not contain corresponding fields. Because of this, the THRU clause is used to define the linkage between PART-INV and CUST-MST. However, in this example, you must define the navigation from CUST-MST through CUST-ORD (adjacent levels), through ORD-LI (also at adjacent levels), and then through PART-REQ (corresponding fields) to get to PART-INV.
PART IV - PRACTICE PROBLEMS

In the two practice problems which follow, you are to write the request and then show the results of the request.

**English Statement of Practice Problem 1** - What is the inventory balance for component number 852? Show the component description and the inventory balance.

**Information Retrieval Request 1** - (Check with answer 3)

Results:

**English Statement of Practice Problem 2** - What parts are used in the construction of components that have a sales price of less than fifty dollars? Show only the part description for these parts.

**Information Retrieval Request 2** - (Check with answer 4)

Results:
ANSWERS TO PRACTICE PROBLEMS

ANSWER 1 -

<table>
<thead>
<tr>
<th>COMP-DESC</th>
<th>EXP-QTY</th>
</tr>
</thead>
<tbody>
<tr>
<td>OILTIGHT SELECTOR SWITCH (WL)</td>
<td>1</td>
</tr>
<tr>
<td>TIME SWITCH (SPST) - 120V</td>
<td>4</td>
</tr>
</tbody>
</table>

ANSWER 2 -

<table>
<thead>
<tr>
<th>CUST-NAME</th>
<th>ORD-QTY</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOWING A/C CO.</td>
<td>112</td>
</tr>
</tbody>
</table>

ANSWER 3 - LIST COMP-DESC and FG-QTY in FG-INV when COMP-NO in FG-INV = 852.

RESULTS:

<table>
<thead>
<tr>
<th>COMP-DESC</th>
<th>FG-QTY</th>
</tr>
</thead>
<tbody>
<tr>
<td>OILTIGHT SELECTOR SWITCH (WL)</td>
<td>492</td>
</tr>
</tbody>
</table>

ANSWER 4 - LIST PART-DESC in PART-INV when FG-UNIT-PR in FG-INV < 50.00, THRU MAT-EXP.

RESULTS:

<table>
<thead>
<tr>
<th>PART-DESC</th>
</tr>
</thead>
<tbody>
<tr>
<td>COIL 12</td>
</tr>
<tr>
<td>BEARING PIN 31</td>
</tr>
</tbody>
</table>
INTRODUCTION

You have been asked to participate in an information retrieval exercise designed to evaluate the performance of managers in a business environment. The exercise is presented in two booklets. The first booklet contains (1) a very brief introduction to the business environment, (2) a description of the data available to management, (3) a description of the rules for retrieving information, and (4) two practice problems. One hour will be allowed to study Booklet One and ask questions. All questions will be resolved during the one-hour study period ONLY.

In the second booklet you will be asked to write instructions to retrieve specific information. When writing the instructions in Booklet Two you will be allowed to refer back to this booklet which contains the procedure for retrieving information and examples. Therefore, while studying this material, emphasis should be on understanding the data and how to retrieve information rather than on memorizing any of the material.

PART I - INTRODUCTION TO SPECIAL ELECTRIC COMPANY

The company assembles and markets select electrical components at a reasonable profit. The components to be manufactured are selected because of their wide variety of uses in industry. Because of this wide usage, the company can produce large quantities during a given production run and maintain a large finished-goods inventory to provide prompt customer deliveries.
The managerial decisions in this exercise involve four major entities. Figure 1 illustrates these entities and the relationships among them.

**FIGURE 1 - MANAGEMENT VIEW OF SALES RELATED DATA**

The entities defined in this view are connected with lines representing the relationships among the entities. The relationship between "customers" and "customer orders" (number 1) is one-to-many. This implies that a customer may place any number of orders, but that a given order will pertain to only one customer. The relationship between "customer orders" and "finished-goods inventory" (number 2), like that between "parts inventory" and "finished-goods inventory" (number 3), is many-to-many. This implies that a customer order may contain any number of components from finished-goods inventory (different
line items) and that a given component in inventory may appear on any number of different orders. Relationship number 3 implies that a given component in finished-goods inventory is assembled from any number of different parts and that a given part may be used in the construction of any number of different electrical components.

PART II - DATA DESCRIPTION

This part provides a detailed description of the data used for decision making by sales management. The data pertaining to the four entities in the management view (Figure 1) are defined in six record types. Figure 2 identifies the record type names and the field names within each record type. Figure 3 provides an example of the kind of data found in each record type. For example, a customer master record (CUST-MST) will contain information about one customer, while a record in finished-goods inventory (FG-INV) will contain information about the inventory on one electrical component.

During the exercise, you will need to refer to Figures 2 and 3 frequently. To make this referral easier these figures are provided as separate handouts. Please refer to these handouts when studying the general descriptions of the record and field names which follow:

1 - The Customer Master record type contains fields of customer number, customer name, salesman code, current year sales, and prior year sales.

2 - The Customer Orders record type contains fields of order number, date the order was placed, and the date the components are to be shipped. This record type contains header information for the customer order, with the individual line items being stored in the next record type.
FIGURE 2 -- EXTERNAL VIEW (NETWORK MODEL)

CUST-MST

CUST-NO  CUST-NAME  SM-CODE  C-YR-SALES  P-YR-SALES

CUST-ORD

ORD-NO  ORD-DATE  SHIP-DATE

ORD-LI

LINE-NO  ORD-UNIT-PR  ORD-QTY

FG-INV

COMP-NO  COMP-DESC  FG-REORD-PT  FG-UNIT-PR  FG-QTY

MAT-EXP

EXP-QTY

PART-INV

PART-NO  PART-DESC  PART-REORD-PT  PART-UNIT-PR  PART-QTY
FIGURE 3 -- SAMPLE NETWORK

CUST-MST

524 GENERAL ELECTRIC CORP. 07 1.243,721 345,722
514 TOWING A/C GO. 09 563,421 420,621

CUST-ORD

147 13-12-67 13-14-67
116 10-10-67 10-12-67
147 13-12-67 13-14-67

ORD-LI

01 113,267 123
02 131,430 298
03 312,130 232
04 312,130 232
05 312,130 232
06 312,130 232
07 312,130 232

PG-INV

651 OILIGHT SELECTOR SWITCH (STD) 4907 5.10 1/4
642 OILIGHT SELECTOR SWITCH (MR) 820 95,10 222
850 TIME SWITCH (SPST) 120V 550 126,07 321

MAT-EXP

1 5
2 7
3 1
4 4
5 4

FAPT-INV

638 CHI 12 80 13,20 127
641 REARING PINS 92 99 3,46 1,91
614 COPPER WIRE 89 1000 .15 1205
652 BACK STOP 121 1200 12,27 1,91
674 SPRING 311 300 .15 .43
682 RG ASCATTER 35 60,33 .1
3. The Order Line Item record type contains fields of line number, sales price per component, and the number of components ordered.

4. The Finished-Goods Inventory record type contains fields of component number, component description, the inventory level at which additional production should be ordered (finished-goods reorder point), sales price per component, and quantity on hand.

5. The Materials Explosion record type with a quantity field only. This quantity represents the quantity of this specific part required to assemble a given electrical component (explosion quantity).

6. The Parts Inventory record type contains fields of part number, part description, part reorder point, purchase price per part, and quantity on hand.

The exercise in Booklet Two will require you to have a thorough understanding of the different record types, their contents, and interrelationships. The interrelationship between adjacent records is defined with a connecting line. The interrelationship between non-adjacent records is defined with two connecting lines and an intervening record type(s). For example, the interrelationship between CUST-MST and ORD-LI is through the CUST-ORD record type.

Due to space limitations, Figure 3 shows only a small portion of the actual data that exists for each record type. However, with these examples, Figure 3 can be used to gain an understanding of the existing relationships between record types. The next two examples illustrate the interrelationships among the record types in the system. Please trace these examples carefully, using Figure 3.

Example 1

1. Looking at the CUST-MST and CUST-ORD record types, we can see that Bowin A/C Co., customer master record number 516,
has two outstanding orders, order numbers 116 and 161.

2 - Following order number 161 to the order line item records, we see this order has two line items, numbers 01 and 02.

3 - Line item 01 is for eighty-three units. Following this record to FG-INV, we can see that the component ordered is number 851, Oiltight Selector Switches (STD), and that the inventory balance is 317.

4 - Looking at the MAT-EXP record type, we see that this component is assembled using three of part number 638 and five of part number 644. In the PART-INV record type, part number 644 is described as "BEARING PIN 31," with a quantity of 30 on-hand.

Example 2

1 - From materials explosion and parts inventory, we can see that part number 614, COPPER WIRE 40, is used in the construction of both component numbers 852 and 850. To construct component number 850 requires four of part number 614.

2 - Looking at ORD-LI, we can see that there are three outstanding orders for component 850. These are line 02 of order 161, line 02 of order 116, and line 03 of order 197. From the CUST-ORD record type, we see order number 197 is for customer number 524.

3 - Looking at CUST-MST, we see that 524 is General Electric Corp.

The purpose of these examples is to illustrate how the lines connecting different sets of records are used to represent the interrelationships.

PART III - RULES FOR WRITING INFORMATION RETRIEVAL REQUESTS

This part is divided into two sections. The first section contains general requirements which must be met when writing requests and definitions of key terms used in the command. The second section describes the command that may be used in a request, along with examples of how this command can be used.
Use Figure 3 as an aid to understanding the examples.

GENERAL REQUIREMENTS AND DEFINITIONS

1 - Precise spelling must be used for all field names and record names (see Figure 2 for these spellings).

2 - A reference to a specific field must contain both the field name and the record name.

   Example 1 - CUST-NAME in CUST-MST

   Example 2 - CUST-NO, CUST-NAME, and SM-CODE in CUST-MST

The second example illustrates that the record type name need not be repeated when referencing multiple fields within the same record type.

3 - A condition may be defined in the LIST command by using the "when" clause. The format of a condition is as follows:

   when reference 1 \{< | \>| =\} reference 2 \{or a value\}.

   Compound conditions may be defined using the "and" and "or" conjunctions.

   Example 1 - when C-YR-SALES in CUST-MST < 500,000

   Example 2 - when P-YR-SALES in CUST-MST > C-YR-SALES in CUST-MST and C-YR-SALES in CUST-MST > 1,000,000

DESCRIPTION OF LIST COMMAND

As an aid to understanding the purpose of this command, try to visualize the results of using the command while studying the following examples.

LIST - This command provides the capability of printing lines using specific fields from one or more record types (see example 1 below). The line or lines to be printed may be based upon an optional condition through the use of the "when" clause (see example 2 below). In addition, you must use one or more "THRU" clauses at the end of the
list command when relating to two record types that are not directly connected (see example 3 below).

Example 1 - LIST COMP-NO and COMP-DESC in FG-INV.

This example will print the component number and description for all finished-goods inventory records.

RESULTS: 
<table>
<thead>
<tr>
<th>COMP-NO</th>
<th>COMP-DESC</th>
</tr>
</thead>
<tbody>
<tr>
<td>851</td>
<td>OILTIGHT SELECTOR SWITCH (STD)</td>
</tr>
<tr>
<td>852</td>
<td>OILTIGHT SELECTOR SWITCH (WL)</td>
</tr>
<tr>
<td>850</td>
<td>TIME SWITCH (SPST) - 120V</td>
</tr>
</tbody>
</table>

Example 2 - LIST COMP-DESC in FG-INV and EXP-QTY in MAT-EXP when PART-NO in PART-INV = 614.

In this example the system will search PART-INV until the condition in the "when" clause is met. It will then print the EXP-QTY in MAT-EXP which is linked to this record, and the COMP-DESC in FG-INV which is linked to the MAT-EXP record. This request does not need a THRU clause because the three record types specified in the request are linked directly.

Write the results of this request in the space which follows:

Check this answer with answer 1 on the last page.

Example 3 - LIST CUST-NAME in CUST-MST and ORD-QTY in ORD-LI when ORD-QTY in ORD-LI > 100 THRU CUST-ORD.
In this example the system will search ORD-LI until the condition in the "when" clause is met. It will then link to the corresponding record in CUST-ORD (because of the THRU clause). From here it will link to the CUST-MST record to print the CUST-NAME. It is important to note that the direction of the linkage is from the record type specified in the "when" clause to the information being printed!

Write the results of this request in the space which follows:

Check this answer with answer 2 on the last page.

Following are four examples of complete information retrieval requests. Pay particular attention to the translation of the English statement of the problem into the request. Similar translations will be required in the writing exercise contained in Booklet Two.

------------------------- EXAMPLE 1 -------------------------

**English Statement of Problem:** Which finished-goods inventory items are priced less than 100 dollars?

**Information Retrieval Request:** List COMP-NO and COMP-DESC in FG-INV when FG-UNIT-PR in FG-INV < 100.00.

The results from this request are:

<table>
<thead>
<tr>
<th>COMP-NO</th>
<th>COMP-DESC</th>
</tr>
</thead>
<tbody>
<tr>
<td>851</td>
<td>OILTIGHT SELECTOR SWITCH (STD)</td>
</tr>
</tbody>
</table>
This request requires the use of only one record type "FG-INV." All requests must be explicit about the record type(s) and field(s) being referenced. Refer to Figure 3 to see how this request for information was satisfied.

------------------------ EXAMPLE 2 ------------------------

English Statement of Problem: What customers are scheduled for shipment on 10-12-83?

Information Retrieval Request: List CUST-NO and CUST-NAME in CUST-MST when SHIP-DATE in CUST-ORD = 10-12-83.

Results:

<table>
<thead>
<tr>
<th>CUST-NO</th>
<th>CUST-NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>516</td>
<td>BOWING A/C CO.</td>
</tr>
</tbody>
</table>

This request requires the use of two record types that are linked directly. No THRU clause is required.

------------------------ EXAMPLE 3 ------------------------

English Statement of Problem: Which finished-goods components are assembled using parts that are below the part reorder point?

Information Retrieval Request: LIST COMP-NO and COMP-DESC in FG-INV and PART-NO in PART-INV when PART-QTY in PART-INV < PART-REORD-PT in PART-INV, THRU MAT-EXP.

Results:

<table>
<thead>
<tr>
<th>COMP-NO</th>
<th>COMP-DESC</th>
<th>PART-NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>851</td>
<td>OILTIGHT SELECTOR SWITCH (STD)</td>
<td>644</td>
</tr>
<tr>
<td>852</td>
<td>OILTIGHT SELECTOR SWITCH (WL)</td>
<td>644</td>
</tr>
</tbody>
</table>
This request involves the use of two record types, FG-INV and PART-INV. However, these two record types do not have a direct linkage. Because of this, the THRU clause is used to define the linkage from PART-INV through MAT-EXP to the FG-INV record type. Note that the THRU clause must be used at the end of the request when referencing a second record type that does not contain a direct linkage.

---------------------- EXAMPLE 4 ----------------------

English Statement of Problem: Which parts are used in the construction of components ordered for Bowing A/C Co.?

Information Retrieval Request: LIST PART-NO and PART-DESC in PART-INV when CUST-NAME in CUST-MST = Bowing A/C Co. THRU CUST-ORD, THRU ORD-LI, THRU FG-INV, THRU MAT-EXP.

Results:

<table>
<thead>
<tr>
<th>PART-NO</th>
<th>PART-DESC</th>
</tr>
</thead>
<tbody>
<tr>
<td>652</td>
<td>BACK STOP 111</td>
</tr>
<tr>
<td>634</td>
<td>SPRING 111</td>
</tr>
<tr>
<td>638</td>
<td>COIL 12</td>
</tr>
<tr>
<td>644</td>
<td>BEARING PIN 31</td>
</tr>
<tr>
<td>606</td>
<td>DD ARMATURES</td>
</tr>
<tr>
<td>614</td>
<td>COPPER WIRE 40</td>
</tr>
</tbody>
</table>

This request involves the use of two record types, PART-INV and CUST-MST. Also, as in example 3, these two record types do not have a direct linkage. Because of this, THRU clauses are used to define the linkage between PART-INV and CUST-MST. However, in this example, you must define the navigation from CUST-MST through CUST-ORD, through ORD-LI, through FG-INV, and then
through MAT-EXP to get to PART-INV.
PART IV - PRACTICE PROBLEMS

In the two practice problems which follow, you are to write the request and then show the results of the request.

**English Statement of Practice Problem 1** - What is the inventory balance for component number 852? Show the component description and the inventory balance.

**Information Retrieval Request 1** - (Check with answer 3)

**Results:**

**English Statement of Practice Problem 2** - What parts are used in the construction of components that have a sales price of less than fifty dollars? Show only the part description for these parts.

**Information Retrieval Request 2** - (Check with answer 4)

**Results:**
ANSWERS TO PRACTICE PROBLEMS

ANSWER 1 -

<table>
<thead>
<tr>
<th>COMP-DESC</th>
<th>EXP-QTY</th>
</tr>
</thead>
<tbody>
<tr>
<td>OILTIGHT SELECTOR SWITCH (WL)</td>
<td>1</td>
</tr>
<tr>
<td>TIME SWITCH (SPST) - 120V</td>
<td>4</td>
</tr>
</tbody>
</table>

ANSWER 2 -

<table>
<thead>
<tr>
<th>CUST-NAME</th>
<th>ORD-QTY</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOWING A/C CO.</td>
<td>112</td>
</tr>
</tbody>
</table>

ANSWER 3 - LIST COMP-DESC and FG-QTY in FG-INV when COMP-NO in FG-INV = 852.

RESULTS:

<table>
<thead>
<tr>
<th>COMP-DESC</th>
<th>FG-QTY</th>
</tr>
</thead>
<tbody>
<tr>
<td>OILTIGHT SELECTOR SWITCH (WL)</td>
<td>492</td>
</tr>
</tbody>
</table>

ANSWER 4 - LIST PART-DESC in PART-INV when FG-UNIT-PR in FG-INV < 50.00, THRU MAT-EXP.

RESULTS:

<table>
<thead>
<tr>
<th>PART-DESC</th>
</tr>
</thead>
<tbody>
<tr>
<td>COIL 12</td>
</tr>
<tr>
<td>BEARING PIN 31</td>
</tr>
</tbody>
</table>
APPENDIX D

INFORMATION RETRIEVAL PROBLEMS
INFORMATION RETRIEVAL PROBLEMS

This booklet presents three problems. You are to write the "list" command necessary to retrieve the information requested in the English statement of each problem. Write the list command only, DO NOT show the results of these list commands.

After writing each request, you are to review the request carefully and circle one of the five numbers following the request. This number will indicate your confidence in the accuracy of the request, and will be interpreted as follows: 1 = very sure it's incorrect, 2 = fairly sure incorrect, 3 = 50-50 chance, 4 = fairly sure correct, and 5 = very sure correct.

---------------- PROBLEMS ----------------

English Statement of Problem 1 - Which components are constructed using parts that have an inventory balance of less than 100? Show both the component number and the component description.

Information Retrieval Request 1 -

My confidence in this request = 1 2 3 4 5

incorrect --------- correct
English Statement of Problem 2 - Which components are on order for customers belonging to salesman number nine? Show the component description and the inventory balance.

Information Retrieval Request 2 -

My confidence in this request = 1 2 3 4 5
incorrect --------- correct

English Statement of Problem 3 - Which customers have orders for components that are below their finished-goods reorder point? Show customer name and component number.

Information Retrieval Request 3 -

My confidence in this request = 1 2 3 4 5
incorrect --------- correct
APPENDIX E

SUBJECT BACKGROUND INFORMATION SHEETS
INFORMATION SHEET

Name ________________________________

Age ______

Sex ______

Number of years of business working experience ______

Number of months of business programming experience ______

Number of months of working experience as a computer user ______

Number of computer classes completed ______
APPENDIX F

MYERS-BRIGGS TYPE INDICATOR
MBTI -- DESCRIPTIVE INFORMATION

The purpose for this appendix is to provide the reader with additional descriptive information about the Myers-Briggs Type Indicator. The instrument itself is not included because of a 1976 copyright. The publishing company is Consulting Psychologists Press, Inc, 577 College Avenue, Palto Alto, California 94306.

The instrument is an eight page test consisting of 166 multiple choice questions. The directions on the cover page inform the user that:

There are no "right" or "wrong" answers to these questions. Your answers will help show how you like to look at things and how you like to go about deciding things. Knowing your own preferences and learning about other people's can help you understand where your special strengths are, what kinds of work you might enjoy and be successful doing, and how people with different preferences can relate to each other and be valuable to society.

Read each question carefully and mark your answers on the separate answer sheet. Make no marks on the question booklet. Do not think too long about any question. If you cannot decide on a question, skip it but be careful that the next space you mark on the answer sheet has the same number as the question you are then answering.

A sample question is:

Does following a schedule
(A) appeal to you, or
(B) cramp you?

Answers are recorded on a standard 8-1/2 by 11, machine gradeable form. However, the tests are typically scored manually through the use of scoring templates. The scores
for each subject are recorded on a subject profile sheet, also available from this publisher. The profile sheet is designed to provide the subject with insights into his own personality. The front of this form pictorially presents the four Jungian dimensions (Extraversion/Introversion [E/I], Sensing/Intuition [S/N], Thinking/Feeling [T/F], and Judging/Perceptive [J/P]), and identifies the point upon each dimension indicated by the subject’s score. The scores are then converted to a four letter "type" code. Each letter in the code represents the directional preference for one of the four dimensions. For example the "type" might be equal to ESFJ, which indicates a higher score on the Extroversion portion of the first dimension, the Sensing portion of the second dimension, the Feeling portion of the third dimension, and the Judging portion of the fourth dimension.

Directional preferences for the four dimensions are also defined on the front of the profile sheet. These definitions include:

E - An E for extraversion probably means you relate more easily to the outer world of people than to the inner world of ideas.

I - An I for introversion probably means you relate more easily to the inner world of ideas than to the outer world of people and things.

S - An S for sensing probably means you would rather work with known facts than look for possibilities and relations.

N - An N for intuition probably means you would rather
look for possibilities and relationships than work with known facts.

T - A T for thinking probably means you base your judgments more on impersonal analysis and logic than on personal values.

F - An F for feeling probably means you base your judgments more on personal values than on impersonal analysis and logic.

J - A J for judging attitude probably means you like a planned, decided, orderly way of life better than a flexible, spontaneous way.

P - A P for the perceptive attitude probably means you like a flexible, spontaneous way of life better than a planned, decided, orderly way.

The reverse side of the profile sheet describes each of the sixteen possible combinations of four letter "type" codes. For example the ESFJ type is described as:

Warm-hearted, talkative, popular, conscientious, born cooperators, active committee members. Need harmony and may be good at creating it. Always doing something nice for someone. Work best with encouragement and praise. Little interest in abstract thinking or technical subjects. Main interest is in things that directly and visibly affect people's lives.
BIBLIOGRAPHY

Books


Articles


Harris, L. R., "User Oriented Data Base Query with the ROBOT
Natural Language Query System," International Journal of

Henderson, John C., and Paul C. Nutt, "The Influence of
Decision Style on Decision Making Behavior," Management

Hoffer, Jeffrey A, "An Empirical Investigation into
Individual Differences in Database Models," Proceedings
of the Third International Conference on Information

Kaiser, Kate M., and Robert P. Bostrom, "Personality
Characteristics of MIS Project Teams: An Empirical Study
and Action-Research Design," MIS Quarterly, 1, 1,
(December, 1982), pp. 43-60.

Keen, P. G. W., "Interactive Computer Systems for Managers,"

Keen, Peter G. W., and Gloria S. Bronsema, "Cognitive Style
Research: A Perspective for Integration," Proceedings of
the Second International Conference on Information

Kilmann, Ralph H., and Ian I. Mitroff, "Qualitative versus
Quantitative Analysis for Management Science: Different
Forms for Different Psychological Types," Interfaces, 6,

Kim, Won, "Relational Database Systems," Computing Surveys,
11, 3, (September 1979), pp. 185-211.

Lien, Edmund Y., "On the Equivalence of Database Models,"
Journal of the Association for Computing Machinery, 29,

Mason, R. O., and Mitroff, I. I., "A Program for Research on
Management Information Systems," Management Science, 19,
1, (1973), pp. 475-487.

165


Teorey, Toby J., "The Logical Record Access Approach to


**Unpublished Materials**

