

AN INVESTIGATION OF INFORMATION DISPLAY  
VARIABLES UTILIZING COMPUTER-GENERATED  
GRAPHICS FOR DECISION SUPPORT SYSTEMS

by

Kenneth B. Gaylin

Thesis Submitted to the Faculty of the  
Virginia Polytechnic Institute and State University  
in partial fulfillment of the requirements  
for the degree of  
MASTER OF SCIENCE  
in  
Industrial Engineering and Operations Research

APPROVED:

---

John G. Casali, Chairman

---

Walter W. Wierwille

---

Paul T. Kemmerling

January, 1985  
Blacksburg, Virginia

AN INVESTIGATION OF INFORMATION DISPLAY  
VARIABLES UTILIZING COMPUTER-GENERATED  
GRAPHICS FOR DECISION SUPPORT SYSTEMS

by

Kenneth B. Gaylin

(ABSTRACT)

The effectiveness of selected computer-generated graphics display variables was examined in a mixed-factors factorial experiment using thirty-two subjects. All subjects performed four different graph reading tasks consisting of point-reading, point-comparison, trend-reading, and trend-comparison. In each task, line, point, bar, and three-dimensional bar graphs were investigated under two levels of task complexity, and two levels of coding (color and black-and-white). The effects of these independent variables on measures of task performance errors, time to complete the task, subjective mental workload, and preference ratings were obtained in real-time by a microcomputer control program. Separate MANOVA analyses of these measures for each task indicated significant effects of graph-type for the point-reading task, main effects of complexity and coding for all tasks, and a graph-by-coding interaction for the point-reading, point-comparison, and trend-reading tasks. Subsequent ANOVA analyses showed significance for these effects across several of the dependent measures which are specified in

the thesis. Recommendations are made for selecting the most effective graph and coding combinations for the particular types of graph-interpretation tasks and complexity levels encountered.

## ACKNOWLEDGEMENTS

Extra special thanks go to Dr. John G. Casali, committee chairman, for his encouragement and assistance in all stages of the planning and execution of this study and his guidance in the preparation of this thesis. Dr. Walter W. Wierwille, and Professor Paul T. Kemmerling are also gratefully acknowledged for serving as committee members and for their advice, support, and critical comments regarding this thesis.

The author gratefully acknowledges the Management Systems Laboratory at Virginia Tech for the opportunity to conduct this study, and the valuable assistance that members of their staff have provided. Funding was provided through a research grant from the U.S. Department of Energy. Special thanks go to \_\_\_\_\_, who served as Project Manager, for providing guidance and material support whenever it was needed.

Several members of the Statistics Department at Virginia Tech contributed to this thesis. Dr. and Mrs. Ray Myers provided invaluable help in the experimental design for this study. Dr. Klaus H. Hinkelmann is to be thanked for his assistance in clarifying counterbalancing measures.

\_\_\_\_\_ in Statistical Consulting are also to be thanked for their help with data analysis

procedures.

To my parents and family go my greatest thanks for always believing in me, and their constant support in all of my endeavors.

## TABLE OF CONTENTS

	<u>Page</u>
ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	iv
LIST OF TABLES.....	viii
LIST OF FIGURES.....	xii
LIST OF APPENDICES.....	xv
INTRODUCTION.....	1
Background.....	1
Significance of the Research.....	3
Application of graphical information display.....	3
Graphics information display in decision- support systems.....	4
LITERATURE REVIEW.....	6
Tabular Versus Graphic Presentation.....	6
Graph-Type.....	11
Horizontal Versus Vertical Graph Orientation.....	13
Graph Complexity Factors.....	16
Coding and Labeling Variables.....	17
Cognitive Style of the User.....	19
Summary of Literature Review.....	20
RESEARCH NEEDS.....	23
Selected Experimental (Independent) Variables.....	24
Decision task-type.....	24
Task complexity.....	24
Coding.....	24
Graph-type.....	25
EXPERIMENTAL METHODOLOGY.....	28
Experimental Design.....	28
Task-type.....	28
Graph-type.....	30
Coding.....	32
Task complexity.....	34
Counterbalancing and Experimental Control.....	36
Subjects.....	38
Apparatus.....	39
Hardware.....	39
Software.....	40
Pretesting.....	41
Experimental Procedures.....	41

Screening and informed consent.....	41
Training (subject practice).....	42
Experimental session.....	43
Dependent Measures.....	44
RESULTS.....	46
General Analysis Procedures.....	46
Data Reduction.....	47
Principal Components Analysis.....	49
MANOVA Analysis Using all Dependent Measures.....	59
Point-reading experiment.....	60
Point-comparison experiment.....	62
Trend-reading experiment.....	64
Trend-comparison experiment.....	66
Summary of MANOVA results.....	68
ANOVA Analyses.....	70
Point-reading experiment.....	70
Point-comparison experiment.....	86
Trend-reading experiment.....	104
Trend-comparison experiment.....	122
MANOVA Analysis of Objective Versus Subjective Measures.....	137
Point-reading experiment.....	139
Point-comparison experiment.....	143
Trend-reading experiment.....	149
Trend-comparison experiment.....	154
DISCUSSION AND CONCLUSIONS.....	157
Main Effects of Complexity.....	157
Main Effects of Graph-Type.....	158
Main Effects of Coding.....	161
Graph-by-Coding Interaction.....	164
Graph-by-Complexity-by-Coding Interaction.....	168
Conclusions and Recommendations.....	168
REFERENCES.....	177
APPENDICES.....	180
VITA.....	229

## LIST OF TABLES

<u>Number</u>	<u>Title</u>	<u>Page</u>
1	Principal Components Analysis for the Point- Reading Experiment Subjective Mental Workload Data.....	51
2	Correlation Coefficients for the Point- Reading Experiment Subjective Mental Workload Data.....	52
3	Principal Components Analysis for the Point- Comparison Experiment Subjective Mental Workload Data.....	53
4	Correlation Coefficients for the Point- Comparison Experiment Subjective Mental Workload Data.....	54
5	Principal Components Analysis for the Trend- Reading Experiment Subjective Mental Workload Data.....	55
6	Correlation Coefficients for the Trend- Reading Experiment Subjective Mental Workload Data.....	56
7	Principal Components Analysis for the Trend- Comparison Experiment Subjective Mental Workload Data.....	57
8	Correlation Coefficients for the Trend- Comparison Experiment Subjective Mental Workload Data.....	58
9	MANOVA Summary Table, Point-Reading Experiment, for the Dependent Measures of RMS Error, Time, MWLMEAN, Preference.....	61
10	MANOVA Summary Table, Point-Comparison Experiment, for the Dependent Measures of Percent Error, Time, MWLMEAN, Preference.....	63
11	MANOVA Summary Table, Trend-Reading Experiment, for the Dependent Measures of Percent Error, Time, MWLMEAN, Preference.....	65



12	MANOVA Summary Table, Trend-Comparison Experiment, for the Dependent Measures of Percent Error, Time, MWLMEAN, Preference.....	67
13	Summary Table for MANOVA p-Values Across all Experiments.....	69
14	ANOVA Summary Table, Point-Reading Experiment, for the Dependent Measure of Time.....	71
15	ANOVA Summary Table, Point-Reading Experiment, for the Dependent Measure of RMS Error.....	72
16	ANOVA Summary Table, Point-Reading Experiment, for the Dependent Measure of MWLMEAN.....	73
17	ANOVA Summary Table, Point-Reading Experiment, for the Dependent Measure of Preference.....	74
18	Newman-Keuls Analysis, Point-Reading Experiment, of Task Completion Time for the Main Effects of Graph-Type.....	75
19	Dependent Measure Mean Scores, Point-Reading Experiment, for Levels of Complexity.....	80
20	Newman-Keuls Analysis, Point-Reading Experiment, of RMS Error for the Graph-by-Coding Interaction.....	84
21	ANOVA Summary Table, Point-Comparison Experiment, for the Dependent Measure of Time.....	87
22	ANOVA Summary Table, Point-Comparison Experiment, for the Dependent Measure of Percent Error.....	88
23	ANOVA Summary Table, Point-Comparison Experiment, for the Dependent Measure of MWLMEAN.....	89
24	ANOVA Summary Table, Point-Comparison Experiment, for the Dependent Measure of Preference.....	90
25	Dependent Measure Mean Scores, Point-Comparison Experiment, for Levels of Coding.....	91
26	Dependent Measure Mean Scores, Point- Comparison Experiment, for Levels of Complexity.....	96

27	Newman-Keuls Analysis, Point-Comparison Experiment, of MWLMEAN for the Graph-by- Coding Interaction.....	102
28	ANOVA Summary Table, Trend-Reading Experiment, for the Dependent Measure Time.....	105
29	ANOVA Summary Table, Trend-Reading Experiment, for the Dependent Measure Percent Error.....	106
30	ANOVA Summary Table, Trend-Reading Experiment, for the Dependent Measure MWLMEAN.....	107
31	ANOVA Summary Table, Trend-Reading Experiment, for the Dependent Measure Preference.....	108
32	Dependent Measure Mean Scores, Trend- Reading Experiment, for Levels of Coding....	109
33	Dependent Measure Mean Scores, Trend-Reading Experiment, for Levels of Complexity.....	114
34	Newman-Keuls Analysis, Trend-Reading Experiment, of Percent Error Score for the Graph- by-Coding Interaction.....	120
35	ANOVA Summary Table, Trend-Comparison Experiment, for the Dependent Measure Time.....	123
36	ANOVA Summary Table, Trend-Comparison Experiment, for the Dependent Measure Percent Error.....	124
37	ANOVA Summary Table, Trend-Comparison Experiment, for the Dependent Measure MWLMEAN.....	125
38	ANOVA Summary Table, Trend-Comparison Experiment, for the Dependent Measure Preference.....	126
39	Dependent Measure Mean Scores, Trend-Comparison Experiment, for Levels of Coding.....	127
40	Dependent Measure Mean Scores, Trend-Comparison Experiment, for Levels of Complexity.....	132
41	MANOVA Summary Table, Point-Reading Experiment, for the Dependent Measures of RMS Error, Time.....	141

42	MANOVA Summary Table, Point-Reading Experiment, for the Dependent Measures of MWLMEAN, Preference.....	142
43	MANOVA Summary Table, Point-Comparison Experiment, for the Dependent Measures of Percent Error, Time.....	144
44	MANOVA Summary Table, Point-Comparison Experiment, for the Dependent Measures of MWLMEAN, Preference.....	145
45	Newman-Keuls Analysis, Point-Comparison Experiment, of MWLMEAN for the Main Effects of Graph-Type.....	146
46	Duncan's Multiple Range Analysis, Point- Comparison Experiment, of MWLMEAN for the Main Effects of Graph-Type.....	147
47	MANOVA Summary Table, Trend-Reading Experiment, for the Dependent Measures of Percent Error, Time.....	150
48	Newman-Keuls Analysis, Trend-Reading Experiment, of Percent Error Score for the Graph-by- Coding-by-Complexity Interaction.....	151
49	MANOVA Summary Table, Trend-Reading Experiment, for the Dependent Measures of MWLMEAN, Preference.....	153
50	MANOVA Summary Table, Trend-Comparison Experiment, for the Dependent Measures of Percent Error, Time.....	155
51	MANOVA Summary Table, Trend-Comparison Experiment, for the Dependent Measures of MWLMEAN, Preference.....	156
52	Cell Means for Graph-Type for all Measures and Tasks.....	160
53	Cell Means for Coding for all Measures and Tasks.....	163
54	Cell Means for Combinations of Graph-Type and Coding for all Measures and Tasks.....	166

## LIST OF FIGURES

<u>Number</u>	<u>Title</u>	<u>Page</u>
1	Factorial experimental design for each of the four experiments.....	29
2	Effect of graph-type, point-reading experiment, on task completion time.....	76
3	Effect of coding, point-reading experiment, on preference rating.....	78
4	Effect of complexity, point-reading experiment, on task completion time.....	81
5	Effect of complexity, point-reading experiment, on MWLMEAN rating.....	82
6	Effect of complexity, point-reading experiment, on preference rating.....	83
7	Effect of graph-by-coding interaction, point-reading experiment, on RMS error score.....	85
8	Effect of coding, point-comparison experiment, on task completion time.....	92
9	Effect of coding, point-comparison experiment, on MWLMEAN rating.....	93
10	Effect of coding, point-comparison experiment, on preference rating.....	94
11	Effect of complexity, point-comparison experiment, on task completion time.....	97
12	Effect of complexity, point-comparison experiment, on percent error score.....	98
13	Effect of complexity, point-comparison experiment, on MWLMEAN rating.....	99
14	Effect of complexity, point-comparison experiment, on preference rating.....	100
15	Effect of graph-type-by-coding interaction, point-comparison experiment, on MWLMEAN rating.....	103

16	Effect of coding, trend-reading experiment, on task completion time.....	110
17	Effect of coding, trend-reading experiment, on MWLMEAN rating.....	111
18	Effect of coding, trend-reading experiment, on preference rating.....	112
19	Effect of complexity, trend-reading experiment, on task completion time.....	115
20	Effect of complexity, trend-reading experiment, on percent error score.....	116
21	Effect of complexity, trend-reading experiment, on MWLMEAN rating.....	117
22	Effect of complexity, trend-reading experiment, on preference rating.....	118
23	Effect of graph-type-by-coding interaction, trend-reading experiment, on percent error score.....	121
24	Effect of coding, trend-comparison experiment, on task completion time.....	128
25	Effect of coding, trend-comparison experiment, on MWLMEAN rating.....	129
26	Effect of coding, trend-comparison experiment, on preference rating.....	130
27	Effect of complexity, trend-comparison experiment, on task completion time.....	133
28	Effect of complexity, trend-comparison experiment, on percent error score.....	134
29	Effect of complexity, trend-comparison experiment, on MWLMEAN rating.....	135
30	Effect of complexity, trend-comparison experiment, on preference rating.....	136
31	Effect of graph-type, point-comparison experiment on MWLMEAN rating.....	148

32	Effect of graph-by-coding-by-complexity interaction, trend-reading experiment, on percent error score.....	152
----	--	-----

LIST OF APPENDICES

<u>Number</u>	<u>Title</u>	<u>Page</u>
I	Examples of Experimental Tasks.....	180
II	Examples of Black-and-White Coded Graphs.....	189
III	Examples of Color and Black-and-White Coding.....	198
IV	General Description of the Experiment for Subjects.....	201
V	Subjects' Informed Consent Document.....	204
VI	Specific Instructions to Subjects.....	207
VII	Subjective Mental Workload Rating Scales.....	221
VIII	Preference Rating Scale.....	227

## INTRODUCTION

### Background

Recent advances in both the hardware and software capabilities of computing systems, as well as vast reductions in cost, have promoted increased use of computer-generated graphics. This is especially the case for microcomputer-based systems in the scientific, business, and management fields. Yet very little is known about the proper use of graphics systems for information display purposes. In addition, graphical displays are being used to present data in a variety of decision-making applications, however, there exists little information concerning the types of representation which are most suitable and effective for a particular application. Within-graph variables that pertain to most types of graphs, such as the complexity of the information being represented and the use of different coding techniques, also have not been adequately investigated in the past. The research project described herein addressed several of these issues. The research was aimed at determining the efficacy of different types of graphs, as well as within-graph variables, in facilitating visually-displayed information perception and assimilation for humans involved



in decision-making tasks. Computer-simulated decision tasks were used and performance measures, such as decision accuracy, time to task completion, subjective mental workload, and user preference were obtained. It is hoped that the information provided by this study will eventually aid in the proper selection of graph-types and the specification of detailed formats for the data being represented within these graph types, for the specific kinds of tasks encountered in computer-based decision support systems.

In support of the experiment, a micro-computer (IBM-PC based) software program, which automatically presented different types of graphical information to subjects and subsequently recorded their decision-making performance on several dependent measures was developed. This software may be useful for future follow-on research in computer-aided decision tasks.

### Significance of the Research

Application of graphical information display. One of the most promising, and rapidly-developing forms of data representation and visual information portrayal is via computer-generated graphical displays on VDUs. This "high-tech" data conveyance approach has enjoyed widespread adoption by business, government, scientific, and other professional communities. Applications for computer-based graphical information displays are quite numerous and diverse. Six general categories of application suggested by Machover (1977) include the following:

1. Real-time image generation.
2. Electrical and mechanical computer-aided design.
3. Command, control, and communication.
4. Scientific graphics.
5. Image processing.
6. Management information systems.

Several prominent advantages have resulted in the recent increased use of computer graphics for information portrayal purposes in the above applications. These advantages, according to Burchi (1980), among other authors, include:

1. The availability of a large variety of [software,

hardware, and] displays of all prices and capabilities, including low-cost microprocessor-based cathode ray tube (CRT) systems.

2. The availability of turn-key or complete end-user application packages.
3. An abundance of documentation in the literature about successful applications, payoffs, and implementation techniques.

Graphics information display in decision-support systems. One application area in which the computer display of information is becoming particularly prevalent is in the framework of management information systems. In computer-based decision support systems, a principal requirement is that large amounts of complex, often diverse, information be reduced in some fashion and represented in a clear, concise, and comprehensible form. Computer graphics displays are employed in this vein, due to the computer's capabilities of storage, organization, rapid access, retrieval, and subsequent display of large bodies of data. Personnel functioning in a decision-making capacity may be able to make more rapid, and in some cases more accurate judgements from properly-displayed computer-based information portrayals. Indeed, recent advances in computer graphics technology offer considerable promise for

improving the information sources from which managerial decisions must be made.

Unfortunately, human factors research attention to problems of information display has lagged considerably behind these technological advances. As evidenced in the following literature review, most human engineering research has addressed the more traditional aspects of display design. It is of the utmost importance, particularly in information systems where decisions have critical ramifications, to optimize the graphical display of data to achieve maximal accuracy and utility in information transfer to the user or decision-maker. No matter how complex, comprehensive, or profound the graphical portrayal, it is of no use if accurate information perception and assimilation on the part of the user does not occur.

## LITERATURE REVIEW

Although a number of guidelines have been proposed for the proper use of graphical techniques (Zelazney and Roche, 1972; Lockwood, 1969), a review of the human factors and psychological literature indicates that little research has been done on the information-transfer performance of modern computer-generated graphic displays. There have, however, been a number of reported studies which have addressed isolated graph-types, using either hard-copy or video display unit (VDU) output as the presentation medium. Much of this body of research is pertinent to the proposed study. The following literature review is organized by the type of display format used in each study.

### Tabular Versus Graphic Presentation

Benbasat and Schroeder (1977) investigated graphical versus tabular (in a table) output formats in an inventory management simulation. Graduate business students acted as inventory control managers whose objectives were to have the least total cost at the end of an unspecified number of simulations. Subjects who received graphical output and graphical decision aids produced the lowest costs. Those who viewed graphical output also used the least number of reports. Tabular presentation produced higher costs and

required more reports.

Smith (1975) allowed subjects to create output displays from information contained in fixed reports. He used graduate business students in the role of production managers whose objectives were to minimize plant costs. Although the results were not statistically significant, cost performance and variability of cost performance tended to be better for the group receiving data on the VDU in graphical form as opposed to tabular form.

Lucas (1981) used a simulation exercise of a consumer product import firm. Middle and upper level managers were provided with a history of orders for the last 10 years and allowed up to eight simulations enabling ordering decisions to be tested using the last five years of data. Performance was measured from the single simulation with the least cost out of the eight runs. Subjects were also tested on their knowledge of the data displayed, and were asked to rate the output format on several parameters. In general, there were no consistent benefits attributable to either the graphics or tabular modes. For example, while the VDU tabular group found the simulation output more useful than the two graphics treatments for demand frequencies, they had significantly poorer scores on a test of inventory understanding. Scores on tests of inventory

understanding and probability understanding were higher for the two groups receiving graphical simulation output of cost data as opposed to tabular, and more "enjoyment" of the exercise was reported by one graphics group. Graphics plus tabular output was found to produce better simulation output, but poorer scores on the inventory comprehension test than graphical output alone ( $p < 0.10$ ).

Lucas and Nielson (1980) used a computerized management logistics simulation where subjects competed against four other firms that were controlled by a common algorithm. Each subject was given a number of options and was told to maximize profits by increasing sales and reducing logistics costs. Treatments were crossed with three types of subjects: 1) students in their first and second years in the MBA program, 2) practicing engineers enrolled in a summer industrial engineering program, and 3) senior executives attending a continuing education program. Graphical and tabular data presented on a VDU were compared with tabular information presented on a VDU, or as hardcopy output. Results of the combined graphics treatments across subject-type indicate that there was no clear overall superiority with this format. Industrial engineers and executives, however, did have higher profits and rate of profit increase when presented with data in graphical form.

Tullis (1981) compared VDU-presented color and black-and-white graphics with narrative and structured formats, in the context of a computer-based telephone line testing system. The narrative format used complete words and phrases interspersed with measured electrical values. The structured format separated logical categories of data with measured electrical characteristics presented in a fixed tabular form. The graphics modes presented a schematic of the telephone line in which a box on the left of the display represented the central office and two bars extending to the right indicated the two wires of the telephone line. Good, marginal, and bad conditions were represented in the schematic with completely filled, cross-hatched, or open-spaced bars, respectively, for the black-and-white condition, and with red, yellow, and white bars, respectively, for the color condition. Results indicate that there were no significant differences in accuracy for the different "coding" methods. There were significant differences, however, in time to complete the task. The two graphics conditions were significantly faster than the narrative format. In a subjective assessment of "quality," the formats were ranked as follows (highest to lowest): 1) color graphics, 2) black-and-white graphics, 3) structured, 4) narrative. Comparison of these subjective ratings using Tukey's HSD test indicated significant differences between



formats 1 and 3, 2 and 4, and 1 and 4, in the above ranking. When asked which method they would prefer to work with on a daily basis, seven out of eight subjects listed color graphics as their first choice, and narrative as their last choice. Assignment of choices (by the subjects) to second and third place was not definitive. That is, subjects' preferences tended to vacillate between the black-and-white graphics and structured format.

Aretz and Calhoun (1982) examined the use of four methods of presenting weapon stores management information to pilots in a flight simulator. The formats were alphanumeric, color pictorial, black-and-white pictorial, and alphanumeric plus color pictorial. Alphanumerics were presented in a tabular format on a separate display. The picture forms were presented over a white or grey representation of the underside of an aircraft. Each weapon-type had a unique and characteristic shape. During the simulation, the pilots were required to answer questions pertaining to weapon store type, quantity, drop mode, system status, and hung bomb mode. Mean response times using the black-and-white pictorial format were significantly longer than for the other three displays. There were also differences between the number of errors per format. From least to most errors these were: 1) alphanumeric, 2) alphanumeric plus color graphics, 3) color

graphics, and 4) black-and-white graphics. Subjective responses indicated a significant preference for the alphanumeric/color graphic format.

Scott and Wickens (1983) investigated graphical versus textual formats in a tactical battlefield scenario. Subjects in the textual format were asked to decide whether they would be attacked from the North or South, based upon the displayed diagnostic value of the direction, and the reliability of that value. With the graphical format, the diagnostic and reliability values defined the height and the width of a rectangle respectively, so that the total worth of the values was integrated as an area that could be directly perceived. Statistical analysis of the data indicated that the graphic display gave more accurate judgements than text.

### Graph-Type

A number of studies have been conducted that have compared the effects of different types of graphs on various performance measures. Graph-type in this instance refers to categorizations such as line, bar, point-plot, pie or response surface graphs, as well as other more specialized graphical techniques.

Schutz (1961a) compared line, vertical-bar, and

horizontal-bar graphs. The study was designed to determine which graph-type resulted in superior performance on a task that required subjects to make complex judgements of trend. A set of decision rules for determining the type of trend (up, down, or remaining the same) and probability of trend (90%, 75%, 50%, or 30%) was used to classify treatment presentations into twelve different patterns. There were significant differences in time to complete the task, with the fastest being the line format followed by vertical, and then horizontal bar graphs. A Friedman two-way, nonparametric analysis of variance also indicated that the line graph produced significantly more accurate judgements than either bar graph.

Verhagen (1981) conducted several experiments with graph-type as a variable. In all experiments, subjects were told to detect "out-of-tolerance" conditions which were delineated as being greater than or less than a region on the graph that was demarcated with solid horizontal lines. In the first experiment, students viewed bar graphs and stroke graphs (the end of a bar graph) presented on a rear-projection screen. The overall mean time needed to count the number of deviating out-of-tolerance variables was significantly shorter for the stroke graph. In addition, the stroke graph produced half as many errors as the bar graph.

A second experiment by Verhagen used the same apparatus as previously, but the task required the subject to detect the deviating variables and to remember their code number which was listed along the X-axis. Presentation times were fixed at one or two seconds. The percentage of all missed deviations was found to be significantly less for the stroke graph, however more errors were made in finding the corresponding names of the variables for the stroke graph.

A third Verhagen study compared stroke and T-type graphs presented on a VDU. In this study, subjects were told to call out the code number of the deviating variables as soon as they were detected. The stroke graph produced half as many errors as the T-type graph and time to complete the task was also significantly shorter with the stroke graph.

#### Horizontal Versus Vertical Graph Orientation

Orientation, in graph design, refers to the location of the independent and dependent measures with respect to the X and Y-axis. Graphs are usually vertically oriented, with the dependent measure, such as productivity or pressure, represented as a scaled value along the Y-axis (ordinate), and an independent variable, such as time or

product-type, along the X-axis (abscissa). In this situation, bars in a bar graph would appear vertically elongated, and lines in a line graph would begin at the leftmost portion of the graph and end at the rightmost portion. On a horizontally-oriented graph, the dependent measure would be represented on the X-axis and the independent variable on the Y-axis. In this instance, bar graphs would have horizontally-elongated bars, and line graphs would have lines proceeding from the bottom to the top of the graph.

Koch, Thomas and Guenther (1982) used line graphs in a process control scenario to investigate time scale orientation and directionality. Measures of performance were time to complete the task and accuracy in responding to questions representative of process control task scenarios. Typical tasks included point-reading, identifying intermittent or cyclic transients that deviate from normal, determining trends, and comparing the present behavior of a parameter with past behavior. Although there were no differences in the accuracy scores between horizontal and vertical graph orientations, there was a reliable orientation by axis-directionality interaction for time to complete the task. It was found that the most rapid interpretation of process trend data occurred with the time scale located on the X-axis and progressing away

from the origin. When the time scale was located on the Y-axis, more rapid interpretation occurred with the time scale progressing towards the origin.

Schutz (1961a), previously mentioned, using a task that required subjects to make complex judgements of trend and probability of trend, also investigated the effects of vertical and horizontal bar graphs. Vertical bar graphs were found to be superior both in accuracy and time to complete the task.

Verhagen (1981) used a vertical versus horizontal bar graph orientation in one of his four studies concerning the detection of out-of-tolerance parameters in process supervision. Again, subjects were to count the number of deviating variables in one case, and in another, find and remember their code number within a one or two second presentation time. The data were analyzed with a matched-pair within-subject nonparametric Sign test. The results indicate that the time needed to count the number of deviating variables was significantly shorter for the vertical bar orientation. There was no significant difference between the vertical and horizontal formats for the identification task.

### Graph Complexity Factors

Several studies have investigated graph parameters that can be loosely categorized as relating to type and amount of complexity within a given, graphically-depicted, decision-making task.

Schutz (1961b) used comparison and point-reading tasks to determine the effects of multiple line or multiple graph displays, the number of lines per task (2,3, or 4), and the degree of confusion among lines. For high confusion, lines in the graphs were crossed (intersected) 4, 7, and 10 times for the 2, 3, and 4 line conditions and were never crossed in the low confusion condition. The point-reading task involved finding the specific value of a point on the graph whereas the comparison task required the comparison of two or more points. A number of important results were obtained. Multiple lines on a single graph required less time for the comparison tasks than multiple, single-line graphs with the equivalent information. However, multiple graphs were faster for point reading. Two lines were faster than 3 and 4 lines. Low confusion was faster than high confusion. On an overall basis, point-reading took less time than point-comparison. In the point-reading task, the multiple line display was adversely affected by increased confusion whereas the multiple graph display was

not. Under the comparison task, both types of graphs were adversely affected by the addition of confusion.

Schutz (1961a), previously mentioned, varied the amount of data and the number of missing data points in a task where subjects were asked to classify graphical presentations into one of twelve types according to trend and probability of trend. Graphs had either 6, 12, or 18 points and were missing a total of either zero, one-sixth or one-third points. The results indicate that both increasing the number of points or increasing the number of missing points increased the time to complete the task and decreased accuracy.

#### Coding and Labeling Variables

A number of different labeling and coding techniques have been examined under experimental conditions. Coding techniques are quite numerous, and can vary from color coding of lines and bars to represent different data sets, to the use of black-and-white cross-hatching, dot and dash patterns on lines, and different types of symbols for individual sets of data values on point-plot graphs. A brief overview of past research that has examined the effects of coding technique follows.

Tullis (1981) in his study of a computer-based



telephone line testing system found no significant differences between color and black-and-white coding schemes using a dependent measure of time to complete the task. A subjective rating measure, however, did reveal a preference for color graphics.

Aretz and Calhoun (1982), previously mentioned, in their study of an aircraft weapons storage display system found that color graphics was associated with less time to complete the task and resulted in fewer errors than black-and-white graphics.

Schutz (1961b) varied line graphs by using either color-coding (red, green, yellow, purple), or black-and-white shape coding (+---+, O---O, O●●O●●, Δ●●Δ●●), to distinguish between different entities. A significant advantage in time to complete the task was found in favor of the color-coded graphs. In addition, subjects indicated that they preferred the color format over black-and-white.

Milroy and Poulton (1978) compared three methods of labeling line graphs using point-reading as the experimental task. Three variables with eight data points each were coded with either a hollow sphere, a solid square, or a hollow triangle, and connected by either a solid or a dotted black line. Labeling was performed using one of three methods: 1) label located directly on the

function, 2) label with a key just below the function, or 3) label within a legend below the X-axis. There was a significant difference in time to complete the task for the three labeling methods, with the directly-labeled function being faster than either the key or legend formats.

### Cognitive Style of the User

Lucas and Nielson (1980), in their computerized management logistics simulation, studied the effect of cognitive style as measured by a modified version of the Myers-Briggs test. Subjects were administered the test scale for "sensing" versus "intuition" measurement. A high intuition score was associated with a more heuristic decision-maker whereas a low score indicate the subject was more analytic. No significant differences were found on the management task between these two cognitive styles.

Lucas (1981), used a test developed by Barkin (1974) to classify subjects into analytic versus heuristic groups. Analytics were defined as those more inclined to work with, and rely on, data in decision-making, while heuristics were inclined to consider the entire problem in a more global sense. Results indicate that heuristic decision-makers, using graphical displays of data, produced the best overall performance. Cognitive style also significantly affected an inventory understanding test score with analytics

performing better than heuristics regardless of which graphics treatment they received.

Benbasat and Schroeder (1977) also used the cognitive style categorization created by Barkin (1974). It was found that cognitive style did not affect decision time or task performance. It did, however, interact significantly with the subjects' knowledge of the inventory management field and the amount of information they requested from a database.

#### Summary of Literature Review

It is clear from the preceding literature survey that a variety of graphical display considerations have been studied in previous research efforts. However, in most of the experiments, the variables studied have been limited to only one or two and they have not been fully investigated in an interactive framework. Very recent advances in information display technology have spawned many new questions regarding state-of-the-art VDU display techniques. Nevertheless, from past research, several guarded conclusions about general concepts in graphical information display can be formulated as follows. However, one should exercise caution when utilizing these conclusions, and refer to the original references noted in the literature review for details in each instance.

1. Graphical representations of data can improve performance over tabular or textual presentations of data on certain types of decision tasks.
2. Graphics are preferred (subjectively) over tabular or textual formats.
3. Line graphs are preferred over bar-type graphs for determining trends in data.
4. Stroke graphs are superior to T-type and bar-type graphs for the detection of out-of-tolerance data.
5. Vertical bar graphs appear to be superior to horizontal bar graphs for most tasks.
6. Multiple line (single) graphs are best for comparison tasks whereas multiple graphs are best for point-reading tasks.
7. Irrelevant or missing data can significantly degrade judgement and decision accuracy.
8. Color-coding is helpful for some tasks and is usually user-preferable over black-and-white symbolic codes. Also, color has high user-acceptance.

9. Labels should be located in as close proximity to their associated graph element as possible.
10. User characteristics, such as cognitive style, appear to influence performance on a task-specific basis.

## RESEARCH NEEDS

From the available literature, it appeared that there were few empirically-based guidelines governing the selection of a graphical display technique for a particular decision-making task. Furthermore, the effects of within-graph variables, such as coding techniques, and the amount of data displayed, needed to be more thoroughly investigated. Their influence on the type of task, and interaction with graph format are of great importance, yet have remained largely unexplored using modern computerized display methods. What follows is a description of the variables selected for investigation in this study and some of the reasoning behind their selection.

Selection of graph variables, and levels of those variables was based upon three criteria:

1. Degree of present, and predicted future, real world use.
2. Amount and type of information gained by experimental study.
3. Technical restrictions and capabilities of IBM-PC computing hardware currently available in the Human Factors Laboratory at Virginia Tech.

### Selected Experimental (Independent) Variables

Decision task-type. Of primary importance to the user or creator of graphs is which type of graph to use for a particular set of data. It is improbable that any one graphical technique will be ideally suited for all purposes. Although decisions based on graphically-presented information may involve complex perceptual, communicative, and cognitive processes, it was possible to break these down into a number of detailed core tasks. It was felt that particular types of decisions may involve the use of various levels of one or more of these core tasks.

Task complexity. This refers to the degree of difficulty of the decision or information retrieval process. It has frequently been shown within the human factors literature that differing tasks can often be expected to show wide response variability to different levels of complexity/difficulty. It was felt that while some decision task-types and graph parameters might be relatively insensitive to high levels of complexity, others would perform adequately only at lower levels.

Coding. Two types of coding are predominantly used in real world settings. These are color, and black-and-white symbols or cross-hatching. There are important economic

considerations involved in the decision between these two formats. Hardware capable of displaying color-coded information is typically more than twice as costly as that used to represent data in a black-and-white format. Furthermore, printing equipment, such as color ink jet systems for reproducing hard-copy colored graphs are very expensive and often do not reproduce the same hues, saturations, and brightness of colors as produced on CRT's. Economic differences between these two coding forms escalate even higher if the cost of multiple-copy reproduction is considered. If both formats (black-and-white, and color) are equally effective and/or equally preferred by users, then it may not be necessary to provide the more costly color capability, at least for certain decision task applications.

Graph-type. A review of the literature indicated that few studies have been done to compare the effects of the major types of graphs presently in use. Thus, there are very few objective guidelines for the selection of a particular type of graph for a specific situation. Due to the applied nature of this study, answers to the question of which type of graph to use for a particular decision task were considered to be of paramount importance. In addition, it was felt that the difference between graph-types would account for a large proportion of variability



in subject performance, and thus be among the most important for the practical choice of a display-type.

In this respect, the choice of graph-types for investigation in this study was, to a large degree, based upon a survey of the wide breadth of available "qualities" that make up varieties of computer-generated graphics. General qualities that were observed in a survey of different types of graphs were:

1. Method of point connection:
  - a) Across maximum values of data points (line graph).
  - b) From the bottom of the graph and proceeding vertically upward to each data point (bar graph).
  - c) No connection between data points (point-plot graph).
  
2. Dimension of the point-plotting or point-connection method:
  - a) Thin line.
  - b) Two-dimensional bar.
  - c) One-dimensional point.
  
3. Method used when representing more than one data set:

- a) Data sets do not obscure or hide parts of other data sets (line and point plot graphs).
- b) Data sets overlap in a horizontal direction (side-by-side bar graphs).
- c) Data sets overlap in a vertical direction (stacked bar graphs).
- d) Data sets take on apparent depth into the screen (three-dimensional bar graphs).

Fortunately, large differences in graph qualities usually occurred across basic graph-types, not within them. For example, line graphs connect maximum data values, use a thin line, and do not hide portions of multiple data sets. Bar graphs, however, connect maximum values of data points with the bottom of the graph, use a thick two-dimensional connector, and obscure portions of multiple data sets. Thus, graph-types with similar predominant characteristics, such as stacked and side-by-side bar graphs, were not included in the study to enable concentration on a set of graph-types with more diverse features.

## EXPERIMENTAL METHODOLOGY

### Experimental Design

Three variables were chosen for a four-experiment study, where each individual experiment utilized a particular task-type. A three-way mixed-factors factorial design (Figure 1) constituted the basis for data collection and analysis for each individual experiment. The levels of these variables are discussed in detail below.

Task-type. Task-type was a between-experiment, fixed-effects, variable with four levels: point-reading, point-comparison, trend-reading, and trend-comparison. Examples of each task for the two levels of complexity are given in Appendix I.

a) Point-reading: Point-reading was the process of determining the exact value of a specific data point. This involved finding a specified data point and determining its exact numerical value with respect to a scaled axis.

b) Point-comparison: This referred to the evaluation of two or more points of data to determine greater than, or less than relationships. Specifically, the points of data in question were located, be they on the same line and set

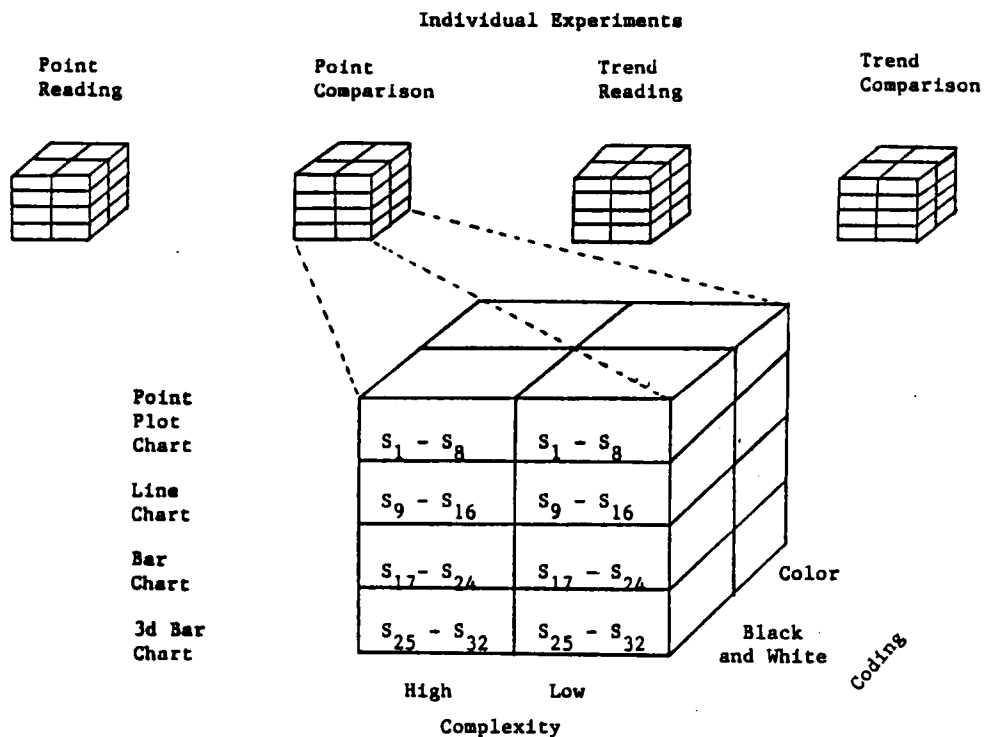


Figure 1. Factorial experimental design for each of the four experiments.

of bars, or different lines or sets of bars, and compared to determine which of the above relationships was true.

c) Trend-reading: This is a task that required the detection of increasing, decreasing, cyclical, or stable tendencies of a set of data points over time. The set of data points within the specified time period would be compared with each other to determine the type of trend with respect to a scaled axis.

d) Trend-comparison: Trend-comparison referred to the discrimination of greater than or less than relationships of the rate of change between two or more sets of data points over time. This was essentially a task that required the determination of the slopes of two or more sets of data over specific time periods, and comparing them to determine the appropriate relationship listed above.

Graph-type. Graph-type was a between-subjects, fixed-effects, variable with four levels: line graph, point-plot graph, bar graph, and three-dimensional bar graph. Listed below are descriptions of the four graph-types used in the study. Examples of black-and-white versions of a sample of the graphs used in the study are given in Appendix II, and examples of color and black-and-white coded graphs are given in Appendix III.

a) Line graphs: These have been, and are likely to

continue to be in wide use. They are available on all commercial graphics packages and were easily introduced into the laboratory setting. Data points were connected by double thickness (two pixel wide) lines to enhance differentiation between lines in both the color-coded, and black-and-white broken-line coded formats. Symbols were not plotted along the line at intersection points, as might be done in some real-world instances, in order to reduce redundancy and derive results for this type of graph in its most basic form. The X and Y-axis on this, and all other graphs, were labeled and had hatch marks delineating numerical or nominal step categories along the axis (refer to Appendix II).

b) Bar graphs: This type of graph is also quite popular and is available on most graphics software. The specific type used was the side-by-side bar graphs. In order to include the desired amount of data points, it was necessary to overlap the bar segments. Thus, some of the segments were partially obscured. This is representative of features of many automated graphics software packages.

c) Point-plot graphs: This is a classic charting technique that continues to be in high demand. This type of graph was different from the other graphs that were investigated in the study because points in the same data

set were distinctly disconnected, in a perceptual sense, from each other. That is, they are not connected by a line, or located in the same row or column as would occur in a line or bar graph, respectively. Points were hollow, and solid geometrical shapes with the presence, or absence of a pixel delineating their respective exact centers.

d) Three-dimensional bar graphs: These types of graphs have not had wide use in the past on micro-computers due to the complexity of the hardware and software that is needed to create them. Many, but certainly not the majority of small computer systems and software packages now have the capability of producing this type of graphic output. One of the features of this type of graph was that variables were represented by width across the screen, and apparent depth into the screen (refer to the three-dimensional bar graphs in Appendix II).

Coding. Coding was a within-subjects, fixed-effects, variable with two levels: color, and black-and-white. Examples of the color, and black-and-white coding conditions are shown in Appendix III. The two categories of this variable are discussed below.

a) Color: In this situation, each separate variable had a color code, and was identified by a key on an empty portion of the graph below the X-axis. In the case of a

line graph, different types of products were depicted by lines of a specific color, or color sequence. For a bar graph, each set of bars would be filled with a different color, or colored cross-hatching. It should be noted that the IBM-PC (on which all experimental tasks were implemented) was capable of displaying only three primary foreground colors. Displaying more than three variables at a time required either the intermixing of pixels, displaying lines with colored sequences of dots and dashes, or the use of colored cross-hatching to fill the solid shapes of bar segments. Of course, this was the practical strategy to take in the development of experimental graphs because it represented what would have to be done in practice with current microcomputer technology.

b) Black-and-white: In this condition, each variable had a different black-and-white pattern. Bars were filled-in with different types and densities of cross-hatching, or with visual texturizing using dot patterns. Data values on point-plot graphs differed by varying unique symbols located where points had been plotted, and line graphs used different line patterns (solid, dotted, dot-dash, dash-dash). Again, refer to Appendix II which presents examples of black-and-white versions of graphs used in the study.



Task complexity. Task complexity was a within-subjects, fixed-effects, variable with two levels: high complexity and low complexity.

a) Low complexity: All low complexity conditions used three data sets per graph. For example, low complexity line graphs had three lines to represent three variables, and bar graphs used three sets of bars. In addition, for three of the four decision tasks (i.e. point-comparison, trend-reading, and trend-comparison), the low complexity condition used easier questions than in the high complexity condition. For point-comparison questions, subjects were asked to identify data points which had either the highest or lowest values of a specified set of points. In the trend-reading experiment, subjects were asked to determine which set of data points was either decreasing or increasing. The trend-comparison experiment required subjects to determine which set of points listed was either increasing or decreasing the most. It was not possible to vary the difficulty level of point-reading questions due to the simple nature of this task.

b) High complexity: For all high complexity tasks, half of the graphs contained five, and the other half contained six data sets per graph. That is, there were five or six sets of lines (for a line graph) or bars (for a

bar graph), depicting, for example different types of products. High complexity tasks also required subjects to answer more difficult questions than in the low complexity condition (again, refer to Appendix I). Point-comparison questions required the subject to determine whether a data value was the second or third highest of the specified set of values. In the trend-reading experiment, subjects were asked to determine which set of points was, for example, increasing, then decreasing, then increasing over a specified time period on the graph. The specified time periods for this task and the trend-comparison task were also greater than in the low complexity condition. Low complexity tasks used time periods across two and three data values (two or three years or months as represented on the X-axis), while high complexity tasks had time periods spanning four and five data values. The high complexity trend-comparison task required subjects to determine which set of points over a specified time period was increasing or decreasing at the second greatest rate, or the third greatest rate in some instances, as compared to other sets of points.

To review and clarify the use of the above variables and levels of those variables in this study it should be understood that each subject (subjects were considered as a random-effects variable) performed all task-types under all

levels of coding and complexity, for a specific graph-type. Within each of the 16 individual cells of the four experimental tasks there were eight subjects who received four presentations each. That is, a subject was asked four different questions for each unique combination of variable levels that comprised a cell in the study. A subject's responses on the four presentations were averaged (arithmetic mean) for a total of eight data points on each dependent measure. All dependent measures were recorded automatically by a control program via keyboard input by subjects.

#### Counterbalancing and Experimental Control

Special consideration was required to control for spurious or confounding effects that might have invalidated the results of this study. Due to the number of questions a subject was asked to answer it was felt that the learning effect could potentially be large. To protect against this effect, the order of question presentation to subjects was completely counterbalanced for the task-type variable by the use of a latin square design. The two variables of complexity and coding could not be completely counterbalanced, but presentation order was alternated to reduce the learning effect. Complexity was alternated on an every other question basis with one-half of the subjects

receiving a low complexity question first, and the other half receiving a high complexity question first. Coding was also alternated, but after every eight questions. Again, one-half of the subjects received eight color-coded graphs first and one-half received eight black-and-white coded graphs first. Since each subject received only one type of graph throughout the experiment, it was not necessary to counterbalance for this effect across presentation order for an individual subject. Each graph-type, however, did have the same question presentation orders to control for a possible spurious main effect across graph-type. It should be emphasized that counterbalancing was done on a per question basis, and not by independent variable. That is, subjects received only one of the four questions in a cell before moving on to the combination of factors comprising the next cell of the counterbalanced order.

Several factors were held constant across graph-types during the development of experimental graphs to reduce confounding. All graphs were constructed using one of four data sets. There were two low complexity and two high complexity data sets. Thus all four graph-types using, for example, low complexity data set number 1 were similar, and could be coupled with the same graph and question pair. This was done to reduce the possibility that a particular

data set was more difficult than another and thus influence a particular set of independent variables. In addition, all graphs used Y-axes that were the same in height, numeric steps, and distance between numeric steps. All labeling of axes was held constant between graph-types, as was the placement of all coded variables in a graph.

Extraneous differences between color and black-and-white coded graphs were held to a minimum by first creating the color version of a graph and subsequently removing chromaticity, via software, to produce the black-and-white version.

### Subjects

Thirty-two subjects were required for the entire experiment. Subjects were paid volunteers from the university community. The subject sample consisted of 16 males and 16 females; four males and four females were used in each graph-type. Each subject was given a visual acuity and color-vision test using a Bausch and Lomb Ortho-Rater and was required to have 22/20 corrected vision (minimum separable acuity) and no color deficiency as defined on the Dvorine Pseudo-Isochromatic Color Vision Test. Subjects who were experienced touch typists on an IBM-PC, who had professional experience with the creation or use of graphical techniques of data representation, or those with

specific academic training in this area were not used in order to reduce the chance of extraneous biases. Assignment of subjects to treatment conditions was random and controlled by a pseudo-random number generator function that was resident in IBM Advanced Basic.

### Apparatus

Hardware. Graphical information was displayed using an IBM-PC microcomputer with two 360 kilobyte disk drives and 256 kilobytes of random access memory (RAM). Subjects viewed the graphs on an IBM Red-Blue-Green color graphics monitor. The screen width of this monitor was 26.5 centimeters long by 19.2 centimeters high. All graphics were presented in the medium resolution color mode, as defined by IBM Advanced Basic, which was capable of displaying 320 horizontal points and 200 vertical points. In this mode, characters were displayed using a 40 column line width. Subject instructions were presented on a Zenith monochrome monitor which used an 80 column line width and was driven by the same IBM-PC as the graph display monitor. The intensity levels of these monitors were pre-set and were not adjustable during the actual experiment.

All input from subjects was accepted through a standard IBM-PC keyboard. Non-essential keys were

deactivated by software to prevent erroneous or accidental input from occurring. Subjects were seated in a standard secretarial chair in front of the IBM-PC and were able to enter information while assuming a typical typing posture. A standard computer terminal table was used to hold the IBM-PC. When seated, the color display was located directly in front of the subject, and the green display was approximately 45 degrees to the left of the subject. The estimated distance from the subject's eyes to each of the display centers ranged from 28 to 45 centimeters. Subjects were allowed to adjust their seating posture and distance to that which was the most comfortable to them throughout the experiment.

Software. All software was run under IBM Disk Operating System Version 2.00 (IBM, 1982). Programming was done in IBM Advanced Basic due to its ability to accommodate moderately complicated quantitative and format-related manipulations. A control program was written to automatically present graphical information and subsequently record decision-making performance on the dependent measures discussed earlier.

Graphical displays were created using several commercially-available software packages, such as the "Energraphics" system. These graphs were then modified to

suit experimental criteria previously mentioned. Graphics thus created were stored on disk to be displayed at the appropriate time by the control program.

### Pretesting

All experimental tasks (decision tasks), graphics display variables, and dependent measures were pretested in a preliminary study with a small number of subjects. The purpose of this pilot study was to target areas in the experimental procedures, and dependent and independent variables, that needed refinement prior to the final experiment.

### Experimental Procedures

Screening and informed consent. Subjects were first asked to read a brief explanation of the experiment (Appendix IV), and a form containing information about subject rights. Signing the "Participants Informed Consent" form (Appendix V) indicated that the subject had read, and agreed to the conditions of the experiment. During this time the experimenter was available to answer any questions the subject had concerning the experiment or subject rights. Once the informed consent form had been signed, the subject was tested for visual acuity and color-vision deficiency using the Bausch and Lomb Ortho-Rater.



Provided that the subject's vision was adequate he or she proceeded to the training phase of the experiment.

Training (subject practice). To reduce the practice effect, subjects were given a sample of 12 graph reading tasks representing the different experimental variables and levels of those variables. Each subject received the same sample and presentation order. For the first eight questions feedback was given in the form of an audible beep if the correct answer had not been chosen. In addition, for the first eight practice questions, the control program did not present the next task until the correct answer was chosen. The last four tasks were presented in the same manner as the experimental tasks. That is, the program did not wait for the input of the correct answer, and no audible beep was given for an incorrect answer. As part of the training period, subjects were first asked to read a specific set of experimental instructions (Appendix VI) describing the format of the study and the use of the computerized rating scales. During this time the experimenter was present in the room to clarify any misunderstanding of the experimental tasks or to answer any questions that the subject had. Subjects must have been able to complete the training tasks in order to continue with the rest of the experiment. No subjects were eliminated from participation for this reason.

Experimental session. Upon completion of training, the subject received a rest break of 5-10 minutes and then began the experimental session. There were 64 separate graph reading tasks (across the four experiments) each taking approximately 1-2 minutes to complete, for a total of about 2.5 hours. A rest break was given after completion of a set of 32 tasks, at the approximate midpoint of the experiment. Subjects were allowed (at their own discretion) to rest briefly to adjust their glasses, go to the water fountain, or ask a question provided that they were between tasks and thus not being timed by the control program. Each task began with the presentation of a question on the monochrome monitor to the subject's left. After the subject read the question, he/she depressed the space bar and a graph, which provided the information necessary to answer the question, appeared on the color monitor located directly in front of the subject. After answering the question, the subject was prompted to rate the task just completed using the six rating scales that were presented on the monochrome monitor. Examples of the rating scales are shown in Appendices VII and VIII. This cycle of events was repeated until the rest break or the end of the experiment. All tasks, as well as rating measures, were self-paced.

### Dependent Measures

Due to the multidimensionality of the information display problem, several dependent measures of both an objective and subjective nature were employed. The selected "objective" measures included task decision errors and time to complete the decision task. Many of the previous studies in this area have used these "objective" measures (e.g., Aretz, 1982; Koch and Edman, 1982; Milroy and Poulton, 1978; Schutz, 1961a; Tullis, 1981; Verhagen, 1981). They were the method of choice if one assumed that the more difficult the task, or the more inadequate the graphical method of data representation, the longer the amount of time that would be needed to determine the correct answer. In addition, it was also assumed that more inadequate methods of graphical data representation would produce larger deviations from a correct answer, and more errors.

Subjective measures of user preference and mental workload were used as indicators of "attention-getting" ability, and perceived mental effort. Measures selected for use included a preference rating scale, and a scale patterned after the multi-descriptor mental workload scales developed by Wierwille and Casali (e.g. cited in Casali, 1982). In previous studies the Mental Workload Multi-

Descriptor Rating Scale was found to show significant sensitivity to load on perceptual tasks involving trained pilots who were performing a simulated flight task (Casali and Wierwille, 1984), but was less sensitive to mediational (Rahimi and Wierwille, 1982), and communicational tasks (Casali and Wierwille, 1983). This scale was also shown to have a distinct monotonic increase in mean ratings across load. The selected subjective mental workload rating scales can be found in Appendix VII.

To determine user preference, a simple bipolar scale (Appendix VIII) was used. The reason for a separate measure of this type was to determine if user preference was significantly different for the different graph-types and within-graph parameters, as well as the other dependent measures. It was felt that a particular type of graph might be preferred, even though certain graph parameters make it very difficult and time-consuming to use.

## RESULTS

### General Analysis Procedures

Data analysis proceeded in several discrete steps. Initially all raw data collected on the IBM-PC were reduced to values applicable to statistical analysis. Due to the multiplicity and expected covariance of the dependent measures, the data were first subjected to a multivariate analysis of variance (MANOVA) procedure. This analysis determined the effects of each independent variable (e.g. graph-type, complexity, and coding) on the composite set of dependent measures (e.g. task time, error score, subjective mental workload mean rating, and preference rating). Given a domain of significance by the MANOVA, the analysis proceeded with individual univariate analysis of variance (ANOVA) procedures on each dependent measure to determine which measures were sensitive to the differences in the independent variables. Sequential multiple-comparison tests (Newman-Keuls Sequential Range test) were then applied, where necessary, to determine the exact location and direction of significant differences between treatment means.

Also, additional MANOVA analyses were conducted to determine if there were differences between the objective

and subjective measures taken in the study.

### Data Reduction

Data reduction procedures were instituted on the IBM-PC, and the reduced data was then transferred to the Virginia Tech IBM 370-165 digital computer. Computations for most statistical procedures were performed using the Statistical Analysis System (SAS) package as implemented on that system (SAS Institute Inc., 1982a).

One procedure performed in the data reduction was the transformation of error scores from discrete, right or wrong, answers to a form approximating a more continuous measure. Based on discussions with the Statistical Consulting Department at Virginia Tech, this data transformation helped to lend credence to the assumption of normality, which fostered the use of parametric statistical analysis with MANOVA procedures. In this procedure the dependent measure scores were averaged across the four questions a subject received in each condition of the study. These means were then each treated as a single observation. This produced error scores between zero and one in steps of 0.25. This procedure was used for all dependent measures in three of the four decision-tasks (point-comparison, trend-reading, and trend-comparison).

In the point-reading task, error scores were different in nature from the other three decision-tasks. The raw data error scores for the point-reading task were measures of exact deviation from the correct answer in the positive (too high) or negative (too low) direction. This being the case, these error scores were in actuality a continuous dependent measure and could have been directly subjected to parametric analyses, however it was decided to treat them in the same manner as in the other three tasks to maintain consistency. Because the use of a simple arithmetic mean of the directional (signed) error scores for the point-reading task might tend to underestimate the actual absolute deviation from the correct answer due to the potential of positive and negative raw data values cancelling each other, it was decided to use the root-mean-square (RMS) error. The formula for RMS error (Friedman, Pisani, Purvis, 1978) is:

$$\text{RMS error} = \sqrt{\frac{\sum (x_i)^2}{N}}$$

where:  $x_i$  = each of four raw data scores  
per cell for a subject

N = total number of scores per cell

RMS error scores for the point-reading task were calculated on the IBM-PC. This data was then transferred to the Virginia Tech IBM 370-165 digital computer. It was also necessary to reduce the subjective mental workload rating scale data, as described in the next section.

### Principal Components Analysis

Due to the similarity in nature of the five subjective mental workload (MWL) measures (i.e. attentional demand, difficulty, complexity, mental workload, and stress level) it was desirable to determine whether they exhibited a significant amount of covariance. If so, this would tend to make MANOVA procedures unreliable and results clouded. To this end, a principal components analysis was conducted. The results, a sample of which is shown in Table 1 for the point-reading experiment, indicated that the first principal component (PRIN1) accounted for a large proportion of the information contained in the five subjective mental workload measures. An examination of the eigenvectors of PRIN1 indicated that essentially similar amounts of information concerning the effect of the independent variables was sampled by each of the five mental workload measures. In these tables, ATTDEM refers to attentional demand rating, DIFF to difficulty rating, COMPLEXR to complexity rating, MENTWORK to mental workload



rating, and STRESS to stress level rating. A simple check of these results can be made by looking at the moderately large correlation coefficients between the MWL measures as displayed in Table 2. Principal components analyses and correlation coefficients for the other three tasks are listed in Tables 3,4,5,6,7, and 8. To reduce the unpredictable effects these measures would have had on an overall MANOVA, the five MWL measures were combined into one arithmetic mean score (MWLMEAN), which was used for all later analyses. This score was, of course, the mean of the ratings obtained on the individual scales. Again, this process of combining the separate mental workload measures into one mean measure was corroborated by the principal components analysis.

TABLE 1

Principal Components Analysis for the Point-Reading  
Experiment Subjective Mental Workload Data

---

	EIGENVALUE	DIFFERENCE	PROPORTION	CUMULATIVE
PRIN1	3.914728	3.522389	0.782946	0.782946
PRIN2	0.392339	0.033898	0.078468	0.861413
PRIN3	0.358441	0.142537	0.071688	0.933102
PRIN4	0.215904	0.097317	0.043181	0.976282
PRIN5	0.118588	.	0.023718	1.000000

EIGENVECTORS


---

	PRIN1	PRIN2	PRIN3	PRIN4	PRIN5
ATTDEM	0.440404	-.684010	-.199406	0.277846	0.470334
DIFF	0.474984	-.260700	-.146507	-.111638	-.820060
COMPLEXR	0.425780	0.633687	-.553596	0.317019	0.100907
MENTWORK	0.461470	0.147351	0.116350	-.809790	0.309896
STRESS	0.431499	0.202223	0.786618	0.392527	-.008329

---

TABLE 2

Correlation Coefficients for the Point-Reading Experiment  
Subjective Mental Workload Data

---

<u>CORRELATIONS</u>					
	ATTDEM	DIFF	COMPLEX	MENTWORK	STRESS
ATTDEM	1.0000	0.8469	0.6282	0.7164	0.6565
DIFF	0.8469	1.0000	0.7385	0.8263	0.7317
COMPLEXR	0.6282	0.7385	1.0000	0.7310	0.6402
MENTWORK	0.7164	0.8263	0.7310	1.0000	0.7551
STRESS	0.6565	0.7317	0.6402	0.7551	1.0000

---

TABLE 3

Principal Components Analysis for the Point-Comparison  
Experiment Subjective Mental Workload Data

	EIGENVALUE	DIFFERENCE	PROPORTION	CUMULATIVE
PRIN1	4.111102	3.588520	0.822220	0.822220
PRIN2	0.522582	0.323885	0.104516	0.926737
PRIN3	0.198697	0.094396	0.039739	0.966476
PRIN4	0.104301	0.040983	0.020860	0.987336
PRIN5	0.063318	.	0.012664	1.000000

<u>EIGENVECTORS</u>					
	PRIN1	PRIN2	PRIN3	PRIN4	PRIN5
ATTDEM	0.454777	-.195992	0.775703	0.100079	0.378199
DIFF	0.476124	-.198520	0.084533	0.013604	-.852388
COMPLEXR	0.456940	-.284251	-.548825	0.576976	0.276218
MENTWORK	0.470854	-.051364	-.299870	-.794769	0.232548
STRESS	0.368763	0.915827	-.002832	0.158872	-.005058

TABLE 4

Correlation Coefficients for the Point-Comparison  
Experiment Subjective Mental Workload Data

---

<u>CORRELATIONS</u>					
	ATTDEM	DIFF	COMPLEXR	MENTWORK	STRESS
ATTDEM	1.0000	0.9033	0.8115	0.8366	0.5968
DIFF	0.9033	1.0000	0.9006	0.9083	0.6273
COMPLEXR	0.8115	0.9006	1.0000	0.8811	0.5665
MENTWORK	0.8366	0.9083	0.8811	1.0000	0.6762
STRESS	0.5968	0.6273	0.5665	0.6762	1.0000

---

TABLE 5

Principal Components Analysis for the Trend-Reading  
Experiment Subjective Mental Workload Data

---

	EIGENVALUE	DIFFERENCE	PROPORTION	CUMULATIVE
PRIN1	4.243793	3.776983	0.848759	0.848759
PRIN2	0.466810	0.293781	0.093362	0.942121
PRIN3	0.173029	0.097233	0.034606	0.976726
PRIN4	0.075796	0.035223	0.015159	0.991886
PRIN5	0.040572	.	0.008114	1.000000

EIGENVECTORS

---

	PRIN1	PRIN2	PRIN3	PRIN4	PRIN5
ATTDEM	0.450213	-.242230	0.788459	-.057160	0.337191
DIFF	0.471164	-.235870	0.022657	0.268980	-.805919
COMPLEXR	0.461957	-.217665	-.502472	0.500021	0.486537
MENTWORK	0.467625	-.053843	-.346052	-.811543	0.008560
STRESS	0.378464	0.914009	0.074758	0.125535	-.002244

---

TABLE 6

Correlation Coefficients for the Trend-Reading Experiment  
Subjective Mental Workload Data

---

<u>CORRELATIONS</u>					
	ATTDEM	DIFF	COMPLEXR	MENTWORK	STRESS
ATTDEM	1.0000	0.9178	0.84 2	0.8560	0.6294
DIFF	0.9178	1.0000	0.9400	0.9228	0.6590
COMPLEXR	0.8432	0.9400	1.0000	0.9217	0.6473
MENTWORK	0.8560	0.9228	0.9217	1.0000	0.7159
STRESS	0.6294	0.6590	0.6473	0.7159	1.0000

---

TABLE 7

Principal Components Analysis for the Trend-Comparison  
Experiment Subjective Mental Workload Data

---

	EIGENVALUE	DIFFERENCE	PROPORTION	CUMULATIVE
PRIN1	4.039990	3.465926	0.807998	0.807998
PRIN2	0.574064	0.356730	0.114813	0.922811
PRIN3	0.217334	0.109844	0.043467	0.966277
PRIN4	0.107490	0.046367	0.021498	0.987775
PRIN5	0.061123	.	0.012225	1.000000

EIGENVECTORS


---

	PRIN1	PRIN2	PRIN3	PRIN4	PRIN5
ATTDEM	0.458369	-.183644	0.733998	-.176953	0.431402
DIFF	0.475799	-.230906	0.171969	0.364319	-.746991
COMPLEXR	0.465519	-.218823	-.512401	0.488181	0.484286
MENTWORK	0.471151	-.067556	-.409534	-.764448	-.146130
STRESS	0.353303	0.927637	0.037419	0.115141	0.003062

---



TABLE 8

Correlation Coefficients for the Trend-Comparison  
Experiment Subjective Mental Workload Data

---

<u>CORRELATIONS</u>					
	ATTDEM	DIFF	COMPLEXR	MENTWORK	STRESS
ATTDEM	1.0000	0.9062	0.8069	0.8250	0.5603
DIFF	0.9062	1.0000	0.9017	0.8760	0.5619
COMPLEXR	0.8069	0.9017	1.0000	0.8957	0.5499
MENTWORK	0.8250	0.8760	0.8957	1.0000	0.6237
STRESS	0.5603	0.5619	0.5499	0.6237	1.0000

---

### MANOVA Analysis Using all Dependent Measures

After data reduction, a multivariate analysis of variance (MANOVA) procedure was used to determine the effects of the independent variables on the set of dependent measures. MANOVA analysis was chosen for three reasons (Finkelstein, Wolf, and Friend, 1977):

1. Prevention of excessive alpha (Type 1) error that would result by the use of separate univariate analysis of variance (ANOVA) procedures.
2. Increased experimental power. Separate ANOVA procedures may fail to detect significance which is spread across more than one dependent measure.
3. The ability to account for the interdependence of dependent measures.

The dependent measures submitted to the initial overall MANOVA analysis were time to complete the task, error, mean of MWL measures (MWLMEAN), and preference. The Wilk's  $\bar{U}$  statistic was used for all effects. All  $\bar{F}$ -ratios and degrees of freedom were converted from the Wilk's  $\bar{U}$  criterion values (SAS Institute Inc., 1982b), and these statistics are reported in the text throughout this section. Each of the four tasks investigated was analyzed as separate and distinct experiments due to the differences

in the types of errors that could occur. For example, the percent error scores collected for the point-comparison, trend-reading, and trend-comparison experiments were comparable in a numerical sense, but were categorically different for each task. That is, one could not equate an error in determining a type of trend (increasing or decreasing) with an error in comparing two points on a graph (determining which point is higher than another). Results of the MANOVA procedures conducted on each experiment are discussed below.

Point-reading experiment. As previously discussed, in the point-reading task, individuals were required to determine the exact value of a point on a graph and RMS error was used for all error scores. All other dependent measures for this task were the same as in the other three tasks. An alpha level of 0.05 was used as a cut-off point to select significant effects. The results of the MANOVA for this task are shown in Table 9. There were four significant effects for this task. These were graph ( $F(12,66)=2.33, p=0.0150$ ), coding ( $F(4,25)=14.70, p=0.0001$ ), complexity ( $F(4,25)=6.26, p=0.0012$ ), and a graph-by-coding interaction ( $F(12,66)=3.47, p=0.0005$ ).

TABLE 9

MANOVA Summary Table, Point-Reading Experiment for the Dependent Measures of RMS Error, Time, MWLMEAN, Preference

<u>Source</u>	<u>dy</u>	$\frac{df}{H}$	$\frac{df}{E}$	<u>U</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>						
Graph (G)	4	3	28	0.3954	2.33	0.0150 *
Subject (S)/G	(Error Term for G)					
<u>Within-Subject</u>						
Coding (CO)	4	1	28	0.2982	14.70	0.0001 *
G X CO	4	3	28	0.2762	3.47	0.0005 *
CO X S/G	(Error Term for CO, G X CO)					
Complexity (CM)	4	1	28	0.4997	6.26	0.0012 *
G X CM	4	3	28	0.6482	0.99	0.4718
CM X S/G	(Error Term for CM, G X CM)					
CO X CM	4	1	28	0.8916	0.76	0.5616
G X CO X CM	4	3	28	0.8497	0.35	0.9753
CO X CM X S/G	(Error Term for CO X CM, G X CO X CM)					

\* significant at  $\alpha = 0.05$

where: dy = number of dependent measures

$\frac{df}{H}$  = degrees of freedom for treatment effect

$\frac{df}{E}$  = degrees of freedom for error effect

U = Wilk's likelihood ratio statistic

$$\frac{|E|}{|E + H|}, \text{ where:}$$

$|E|$  = determinant of sum of squares and cross-products for error

$|E + H|$  = determinant of the sum of the sum of squares and cross-products matrix for error, and the sum of squares and cross-products matrix for treatment

Point-comparison experiment. In this task subjects were asked to determine greater than, less than, and equality relationships between two or more points. Error scores in this and the following two tasks were in the form of percent error scores. The results for this task, shown in Table 10, indicate significant effects for coding ( $F(4,25)=18.13$ ,  $p=0.0001$ ), complexity ( $F(4,25)=67.13$ ,  $p=0.0001$ ), and the graph-by-coding interaction ( $F(12,66)=2.13$ ,  $p=0.0262$ ).

TABLE 10

MANOVA Summary Table, Point-Comparison Experiment, for the Dependent Measures of Percent Error, Time, MWLMEAN, Preference

<u>Source</u>	<u>dy</u>	$\frac{df}{H}$	$\frac{df}{E}$	<u>U</u>	<u>F</u>	<u>p</u>
Graph (G)	4	3	28	0.4995	1.66	0.0965
Subject (S)/G	(Error Term for G)					
<u>Within-Subject</u>						
Coding (CO)	4	1	28	0.2563	18.13	0.0001 *
G X CO	4	3	28	0.0262	2.13	0.0262 *
CO X S/G	(Error Term for CO, G X CO)					
Complexity (CM)	4	1	28	0.0851	67.13	0.0001 *
G X CM	4	3	28	0.7120	0.76	0.6898
CM X S/G	(Error Term for CM, G X CM)					
CO X CM	4	1	28	0.9288	0.48	0.7512
G X CO X CM	4	3	28	0.7524	0.63	0.8105
CO X CM X S/G	(Error Term for CO X CM, G X CO X CM)					

\* significant at  $\alpha = 0.05$

where: dy = number of dependent measures

$\frac{df}{H}$  = degrees of freedom for treatment effect

$\frac{df}{E}$  = degrees of freedom for error effect

U = Wilk's likelihood ratio statistic

$$\frac{|E|}{|E + H|}, \text{ where:}$$

$|E|$  = determinant of sum of squares and cross-products for error

$|E + H|$  = determinant of the sum of the sum of squares and cross-products matrix for error, and the sum of squares and cross-products matrix for treatment

Trend-reading experiment. This task required the detection and classification of any increasing or decreasing tendencies of a set of data points over time. The results of the MANOVA for this task are displayed in Table 11 and are similar in nature to the previous task. That is, significant effects were found for coding ( $F(4,25)=19.95, p=0.0001$ ), complexity ( $F(4,25)=93.79, p=0.0001$ ), and a graph-by-coding interaction ( $F(12,66)=2.19, p=0.0224$ ).

TABLE 11

MANOVA Summary Table, Trend-Reading Experiment, for the Dependent Measures of Percent Error, Time, MWLMEAN, Preference

<u>Source</u>	<u>dv</u>	<u>df</u> H	<u>df</u> E	<u>U</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>						
Graph (G)	4	3	28	0.6014	1.17	0.3207
Subject (S)/G	(Error Term for G)					
<u>Within-Subject</u>						
Coding (CO)	4	1	28	0.2385	19.95	0.0001 *
G X CO	4	3	28	0.4146	2.19	0.0224 *
CO X S/G	(Error Term for CO, G X CO)					
Complexity (CM)	4	1	28	0.0624	93.79	0.0001 *
G X CM	4	3	28	0.4923	1.70	0.0868
CM X S/G	(Error Term for CM, G X CM)					
CO X CM	4	1	28	0.9328	0.45	0.7714
G X CO X CM	4	3	28	0.5247	1.53	0.1366
CO X CM X S/G	(Error Term for CO X CM, G X CO X CM)					

\* significant at  $\alpha = 0.05$

where: dv = number of dependent measures

$\frac{df}{H}$  = degrees of freedom for treatment effect

$\frac{df}{E}$  = degrees of freedom for error effect

U = Wilk's likelihood ratio statistic

$$\frac{|E|}{|E + H|}, \text{ where:}$$

|E| = determinant of sum of squares and cross-products for error

|E + H| = determinant of the sum of the sum of squares and cross-products matrix for error, and the sum of squares and cross-products matrix for treatment



Trend-comparison experiment. Trend-comparison tasks required the discrimination of greater than, less than, and equality relationships of the rate of change between two or more sets of data points over time. Results from the MANOVA, shown in Table 12, demonstrated significance only for the variables of coding ( $F(4,25)=20.60, p=0.0001$ ), and complexity ( $F(4,25)=103.48, p=0.0001$ ).

TABLE 12

MANOVA Summary Table, Trend-Comparison Experiment, for the Dependent Measures of Percent Error, Time, MWLMEAN, Preference

<u>Source</u>	<u>dv</u>	<u>df</u> H	<u>df</u> E	<u>U</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>						
Graph (G)	4	3	28	0.5433	1.44	0.1726
Subject (S)/G	(Error Term for G)					
<u>Within-Subject</u>						
Coding (CO)	4	1	28	0.2327	20.60	0.0001 *
G X CO	4	3	28	0.6273	1.07	0.4017
CO X S/G	(Error Term for CO, G X CO)					
Complexity (CM)	4	1	28	0.0569	103.48	0.0001 *
G X CM	4	3	28	0.6073	1.15	0.3384
CM X S/G	(Error Term for CM, G X CM)					
CO X CM	4	1	28	0.8130	1.44	0.2510
G X CO X CM	4	3	28	0.7211	0.73	0.7190
CO X CM X S/G	(Error Term for CO X CM, G X CO X CM)					

\* significant at  $\alpha = 0.05$

where: dv = number of dependent measures

df<sub>H</sub> = degrees of freedom for treatment effect

df<sub>E</sub> = degrees of freedom for error effect

U = Wilk's likelihood ratio statistic

$$\frac{|E|}{|E + H|}, \text{ where:}$$

|E| = determinant of sum of squares and cross-products for error

|E + H| = determinant of the sum of the sum of squares and cross-products matrix for error, and the sum of squares and cross-products matrix for treatment

Summary of MANOVA Results. A summary of the probability values from the MANOVA results is shown in Table 13. When viewed collectively it is apparent that there is a great deal of similarity in the results across tasks. Very strong effects are present for both coding and complexity in all tasks, while a graph-by-coding interaction occurred for only three tasks. A main effect of graph was present only in the point-reading experiment.

TABLE 13

Summary Table for MANOVA p-Values Across all Experiments \*

---

VARIABLE	EXPERIMENT			
	Point Reading	Point Comparison	Trend Reading	Trend Comparison
Graph (G)	0.0150	-----	-----	-----
Coding (CO)	0.0001	0.0001	0.0001	0.0001
Complexity (CM)	0.0012	0.0001	0.0001	0.0001
G X CO	0.0005	0.0262	0.0224	-----
G X CM	-----	-----	-----	-----
CO X CM	-----	-----	-----	-----
G X CO X CM	-----	-----	-----	-----

---

\* Only significant effects,  $p \leq 0.05$  are shown

## ANOVA Analyses

A univariate analysis of variance procedure was performed separately on each dependent measure to determine their sensitivity to changes in the graph, coding, and complexity independent variables. Only effects which were found to be significant in the overall MANOVA were further examined, as dictated by the Hummel-Sligo procedure. The ANOVA, and follow-on analyses will be discussed separately for each decision-task.

Point-reading experiment. ANOVA analysis of the task completion time, RMS error, MWLMEAN, and preference scores for this task are shown in Tables 14, 15, 16, and 17 respectively. The main effect of graph was significant only for the time score ( $F(3,28)=4.79, p=0.0081$ ). Newman-Keuls analysis of the treatment means (Table 18) indicated that task time was nearly equal for all graphs except the three-dimensional bar graph (3DBAR), which took significantly longer. These results are depicted graphically in Figure 2.

TABLE 14

ANOVA Summary Table, Point-Reading Experiment, for the  
Dependent Measure of Time

<u>Source</u>	<u>df</u>	<u>SS</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>				
Graph (G)	3	6361.080	4.79	0.0081 *
Subject (S)/G	28	12390.842		
<u>Within-Subject</u>				
Coding (CO)	1	1.228	0.04	0.8528
G X CO	3	142.814	1.36	0.2755
CO X S/G	28	980.573		
Complexity (CM)	1	92.981	4.27	0.0483 *
G X CM	3	122.209	1.87	0.1578
CM X S/G	28	610.320		
CO X CM	1	11.741	0.25	0.6225
G X CO X CM	3	61.141	0.43	0.7730
CO X CM X S/G	28	1326.612		
<u>Total</u>	<u>127</u>			

\* significant at  $\alpha = 0.05$

TABLE 15

ANOVA Summary Table, Point-Reading Experiment, for the  
Dependent Measure of RMS Error

<u>Source</u>	<u>df</u>	<u>SS</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>				
Graph (G)	3	19.232	2.14	0.1177
Subject (S)/G	28	83.900		
<u>Within-Subject</u>				
Coding (CO)	1	0.197	0.09	0.7661
G X CO	3	22.185	3.39	0.0319 *
CO X S/G	28	61.165		
Complexity (CM)	1	0.001	0.00	0.9829
G X CM	3	5.691	0.49	0.6890
CM X S/G	28	107.420		
CO X CM	1	0.236	0.06	0.8104
G X CO X CM	3	4.002	0.33	0.8028
CO X CM X S/G	28	112.782		
<u>Total</u>	<u>127</u>			

\* significant at  $\alpha = 0.05$

TABLE 16

ANOVA Summary Table, Point-Reading Experiment, for the  
Dependent Measure of MWLMEAN

<u>Source</u>	<u>df</u>	<u>SS</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>				
Graph (G)	3	38.773	2.32	0.0965
Subject (S)/G	28	155.755		
<u>Within-Subject</u>				
Coding (CO)	1	0.765	1.58	0.2194
G X CO	3	1.724	1.18	0.3334
CO X S/G	28	13.587		
Complexity (CM)	1	1.300	9.31	0.0050 *
G X CM	3	0.998	2.38	0.0905
CM X S/G	28	3.911		
CO X CM	1	0.125	1.63	0.2117
G X CO X CM	3	0.077	0.34	0.7988
CO X CM X S/G	28	2.142		
<u>Total</u>	<u>127</u>			

\* significant at  $\alpha = 0.05$



TABLE 17

ANOVA Summary Table, Point-Reading Experiment, for the  
Dependent Measure of Preference

<u>Source</u>	<u>df</u>	<u>SS</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>				
Graph (G)	3	45.048	1.21	0.3260
Subject (S)/G	28	34.818		
<u>Within-Subject</u>				
Coding (CO)	1	160.316	61.11	0.0001 *
G X CO	3	22.060	2.80	0.0581
CO X S/G	28	73.451		
Complexity (CM)	1	17.441	18.48	0.0002 *
G X CM	3	4.645	1.64	0.2024
CM X S/G	28	26.427		
CO X CM	1	0.215	0.70	0.4101
G X CO X CM	3	0.552	0.60	0.6218
CO X CM X S/G	28	8.623		
<u>Total</u>	<u>127</u>			

\* significant at  $\alpha = 0.05$

TABLE 18

Newman-Keuls Analysis, Point-Reading Experiment, of Task Completion Time for the Main Effects of Graph-Type \*

---

GRAPH-TYPE				
GRAPH	BAR	LINE	POINT	3DBAR
MEAN	18.68	19.54	20.28	35.73
TIME				
(sec.)				

---

\* Treatments underlined by a common line do not differ from each other at  $p \leq 0.05$

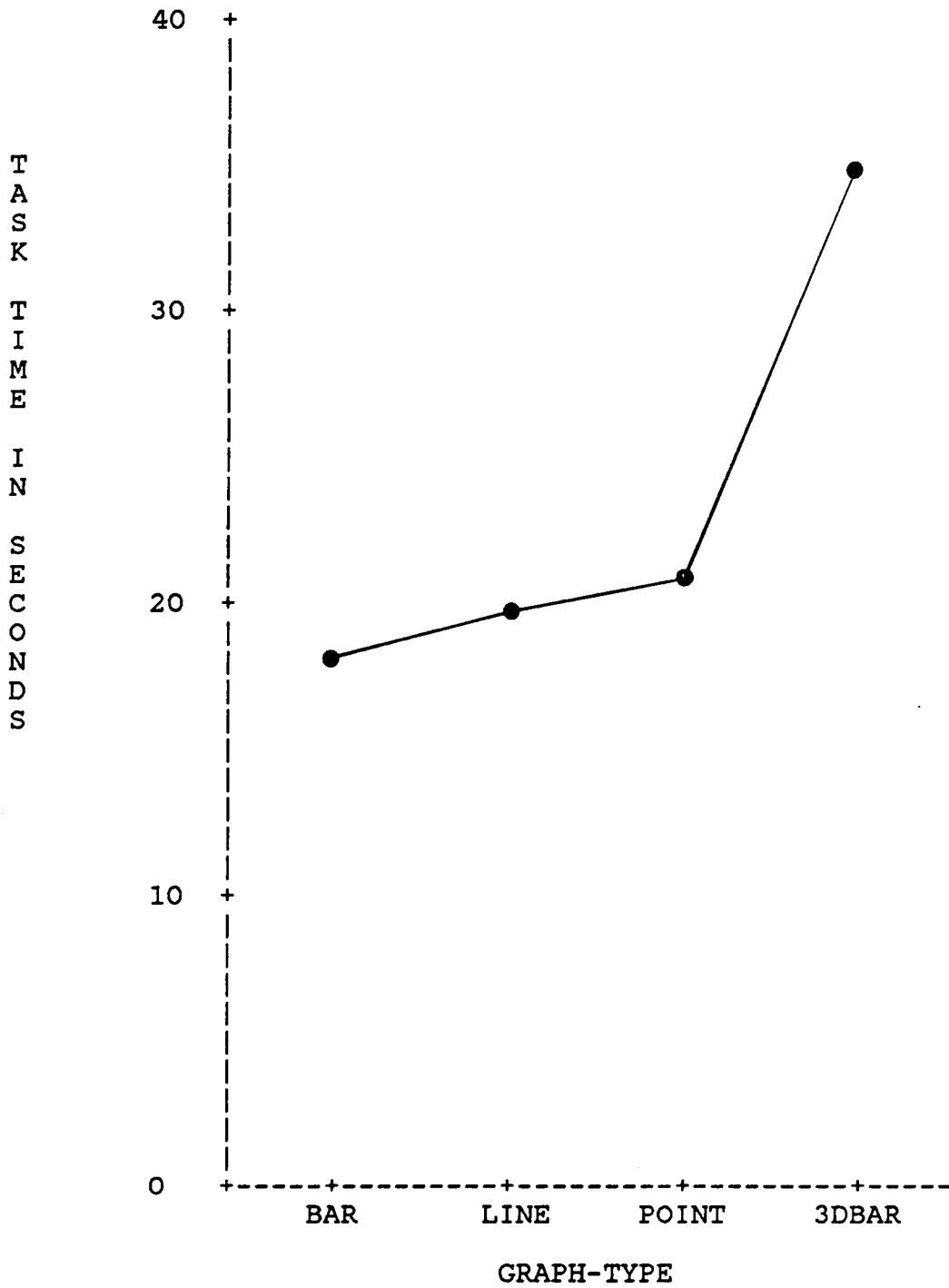


Figure 2. Effect of graph-type, point-reading experiment, on task completion time.

The main effect of coding in the point-reading experiment was significant only with the subjective rating score for preference ( $F(1,28)=61.11, p=0.0001$ ). As illustrated in Figure 3, subjects preferred color over black-and-white coding.

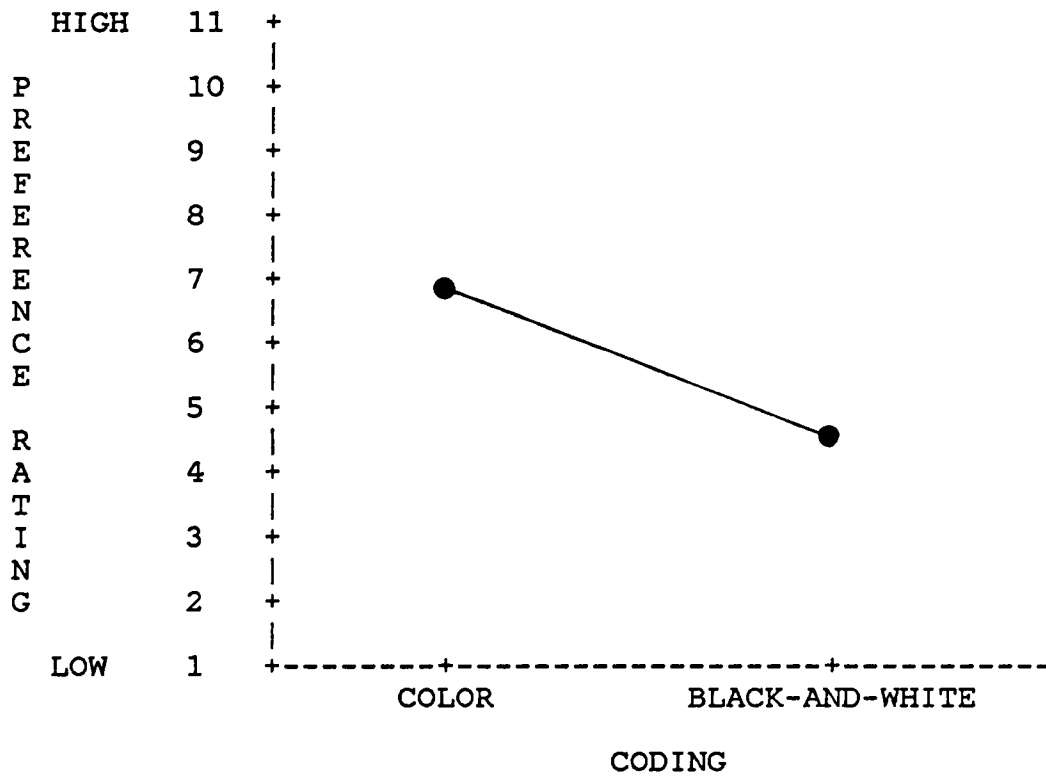


Figure 3. Effect of coding, point-reading experiment, on preference rating.

Complexity effects in the point-reading experiment were present for the dependent measures of time ( $F(1,28)=4.27, p=0.0483$ ), MWLMEAN ( $F(1,28)=9.31, p=0.005$ ), and preference ( $F(1,28)=18.48, p=0.0002$ ). As indicated by the mean scores in Table 19, subjects needed more time to complete high complexity tasks, rated them higher on the MWL measures, and as expected, preferred less complex graphs. These results are depicted graphically in Figures 4, 5, and 6.

The graph-by-coding interaction produced significant results only for the error score dependent measure ( $F(3,28)=3.39, p=0.0319$ ), however preference was nearly significant ( $F(3,28)=2.80, p=0.0581$ ). A Newman-Keuls analysis of error, depicted in Table 20, indicated significant differences in RMS error scores only between black-and-white point-plot graphs and the two conditions of black-and-white bar and line graphs. This is shown graphically in Figure 7.

TABLE 19

Dependent Measure Mean Scores, Point-Reading Experiment,  
for Levels of Complexity

---

	TIME	MWLMEAN	PREFERENCE
HIGH COMPLEXITY	24.41	4.03	5.24
LOW COMPLEXITY	22.70	3.83	5.98

---

where: TIME is in units of seconds

MWLMEAN is an interval scale in units of 1 (low) to 11 (high)

PREFERENCE is an interval scale in units of 1 (low) to 11 (high)

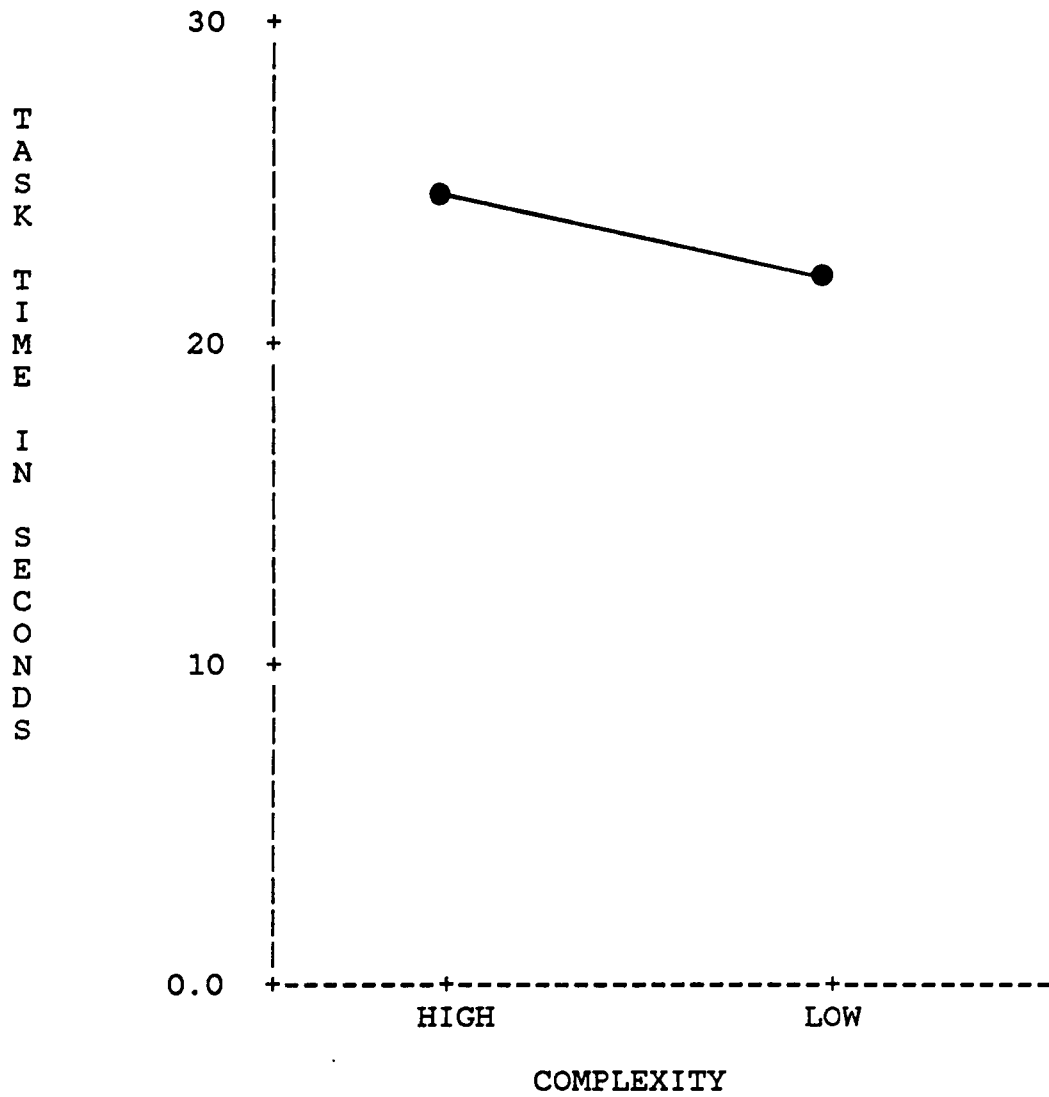


Figure 4. Effect of complexity, point-reading experiment, on task completion time.



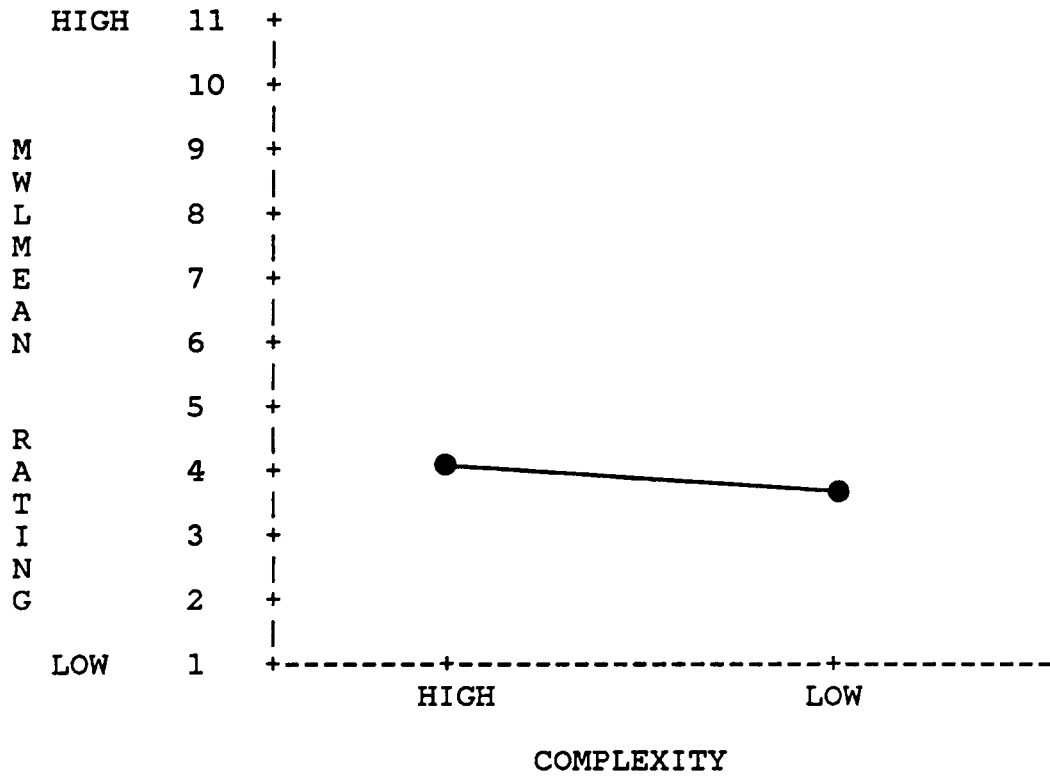


Figure 5. Effect of complexity, point-reading experiment, on MWLMEAN rating.

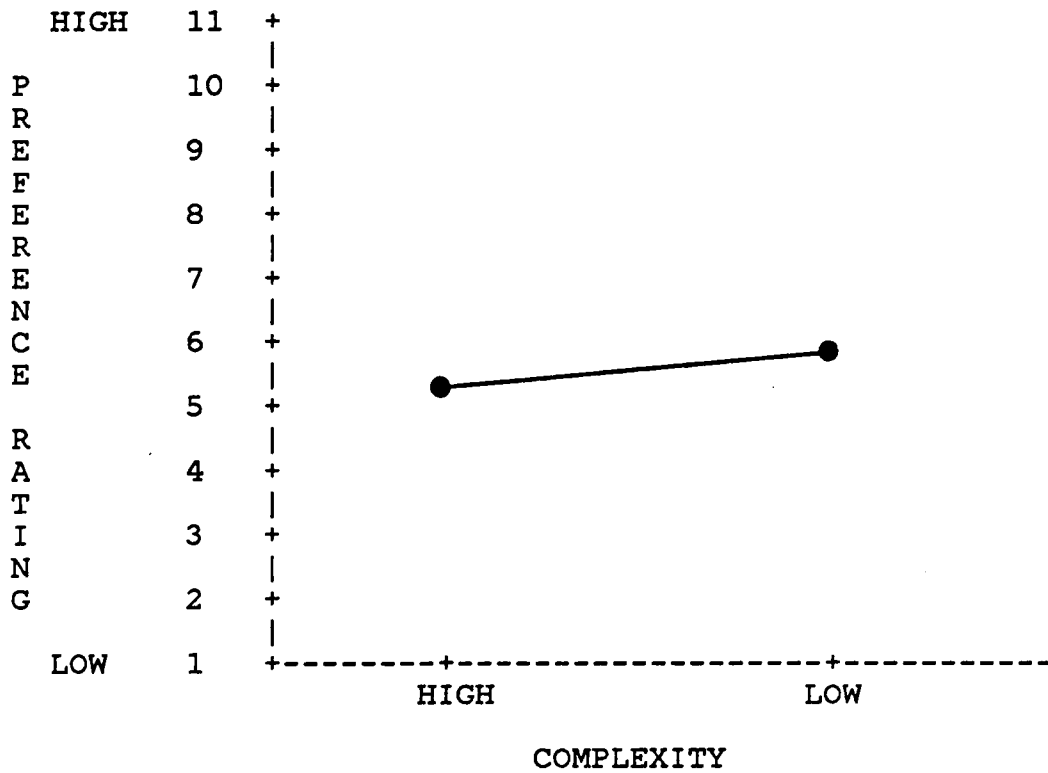


Figure 6. Effect of complexity, point-reading experiment, on preference rating.

TABLE 20

Newman-Keuls Analysis, Point-Reading Experiment, of RMS Error for the Graph-by-Coding Interaction \*

---

GRAPH	BAR	LIN	PNT	BAR	LIN	3DB	3DB	PNT
CODING	B&W	B&W	COL	COL	COL	COL	B&W	B&W
MEAN	.158	.310	.742	.801	.844	.937	.980	2.190
RMS								
ERROR								

---

\* Treatments underlined by a common line do not differ from each other at  $p \leq 0.05$

where:

- BAR = Bar graph
- LIN = Line graph
- PNT = Point-plot graph
- 3DB = Three-dimensional bar graph
- B&W = Black-and-white coding
- COL = Color-coding

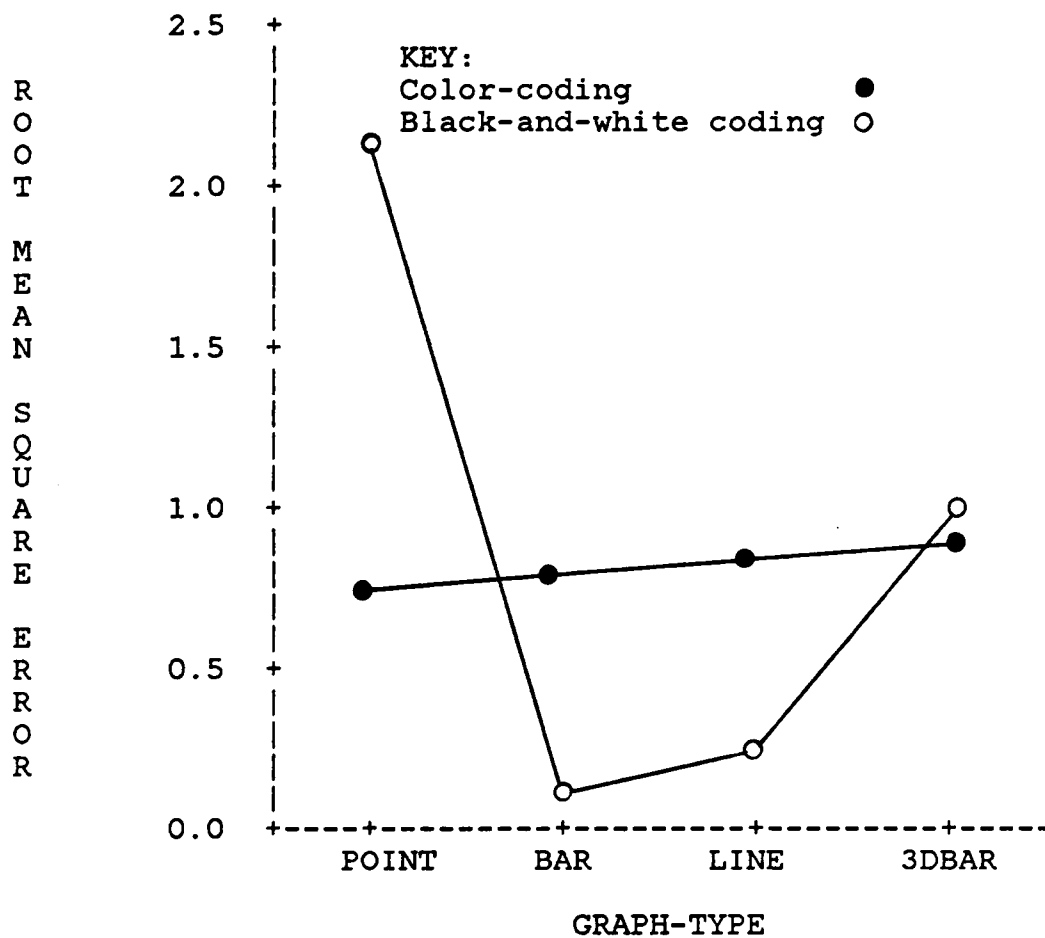


Figure 7. Effect of graph-by-coding interaction, point-reading experiment, on RMS error score.

Point-comparison experiment. ANOVA summary tables of the point-comparison experiment for the four dependent measures of time, percent error, MWLMEAN, and preference scores are shown in Tables 21, 22, 23, and 24 respectively.

The main effect of coding was significant for time ( $F(1,28)=10.82$ ,  $p=0.0027$ ), MWLMEAN ( $F(1,28)=17.17$ ,  $p=0.0003$ ), and preference ( $F(1,28)=10.60$ ,  $p=0.0001$ ). Subjects needed more time to complete the point-comparison tasks when using graphs with black-and-white coding, gave them higher MWLMEAN ratings, and lower preference ratings, than graphs using color coding. Mean values for these results are shown in Table 25, and are depicted graphically in Figures 8, 9, and 10.

TABLE 21

ANOVA Summary Table, Point-Comparison Experiment, for the  
Dependent Measure of Time

<u>Source</u>	<u>df</u>	<u>SS</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>				
Graph (G)	3	5688.665	2.10	0.1234
Subject (S)/G	28	25335.816		
<u>Within-Subject</u>				
Coding (CO)	1	1351.545	10.82	0.0027 *
G X CO	3	568.065	1.52	0.2320
CO X S/G	28	3497.286		
Complexity (CM)	1	44717.021	172.74	0.0001 *
G X CM	3	128.763	0.17	0.9185
CM X S/G	28	7248.139		
CO X CM	1	92.813	0.98	0.3313
G X CO X CM	3	386.266	1.36	0.2764
CO X CM X S/G	28	2658.392		
<u>Total</u>	<u>127</u>			

\* significant at  $\alpha = 0.05$

TABLE 22

ANOVA Summary Table, Point-Comparison Experiment, for the  
Dependent Measure Percent Error

<u>Source</u>	<u>df</u>	<u>SS</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>				
Graph (G)	3	0.257	1.66	0.1991
Subject (S)/G	28	1.453		
<u>Within-Subject</u>				
Coding (CO)	1	0.070	2.10	0.1584
G X CO	3	0.085	0.86	0.4755
CO X S/G	28	0.937		
Complexity (CM)	1	1.125	33.05	0.0001 *
G X CM	3	0.203	1.99	0.1385
CM X S/G	28	0.953		
CO X CM	1	0.000	0.00	1.0000
G X CO X CM	3	0.109	0.92	0.4438
CO X CM X S/G	28	1.109		
<u>Total</u>	<u>127</u>			

\* significant at  $\alpha = 0.05$

TABLE 23

ANOVA Summary Table, Point-Comparison Experiment, for the  
Dependent Measure MWLMEAN

<u>Source</u>	<u>df</u>	<u>SS</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>				
Graph (G)	3	28.314	3.12	0.0417 **
Subject (S)/G	28	84.628		
<u>Within-Subject</u>				
Coding (CO)	1	7.387	17.17	0.0003 *
G X CO	3	4.212	3.26	0.0360 *
CO X S/G	28	12.044		
Complexity (CM)	1	191.467	160.44	0.0001 *
G X CM	3	2.444	0.68	0.5701
CM X S/G	28	33.414		
CO X CM	1	0.134	0.73	0.3995
G X CO X CM	3	0.385	0.70	0.5601
CO X CM X S/G	28	5.146		
<u>Total</u>	<u>127</u>			

\* significant at  $\alpha = 0.05$

\*\* not significant in MANOVA, therefore not considered here



TABLE 24

ANOVA Summary Table, Point-Comparison Experiment, for the  
Dependent Measure Preference

<u>Source</u>	<u>df</u>	<u>SS</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>				
Graph (G)	3	27.423	1.22	0.3221
Subject (S)/G	28	210.433		
<u>Within-Subject</u>				
Coding (CO)	1	192.570	70.60	0.0001 *
G X CO	3	17.402	2.13	0.1193
CO X S/G	28	76.371		
Complexity (CM)	1	146.632	105.50	0.0001 *
G X CM	3	7.730	1.85	0.1604
CM X S/G	28	38.917		
CO X CM	1	0.330	0.48	0.4940
G X CO X CM	3	0.392	0.19	0.9021
CO X CM X S/G	28	19.246		
<u>Total</u>	<u>127</u>			

\* significant at  $\alpha = 0.05$

TABLE 25

Dependent Measure Mean Scores, Point-Comparison Experiment,  
for Levels of Coding

---

	TIME	MWLMEAN	PREFERENCE
BLACK & WHITE CODING	58.52	5.35	4.16
COLOR CODING	52.02	4.87	6.62

---

where: TIME is in units of seconds

MWLMEAN is an interval scale in units of 1 (low) to 11 (high)

PREFERENCE is an interval scale in units of 1 (low) to 11 (high)

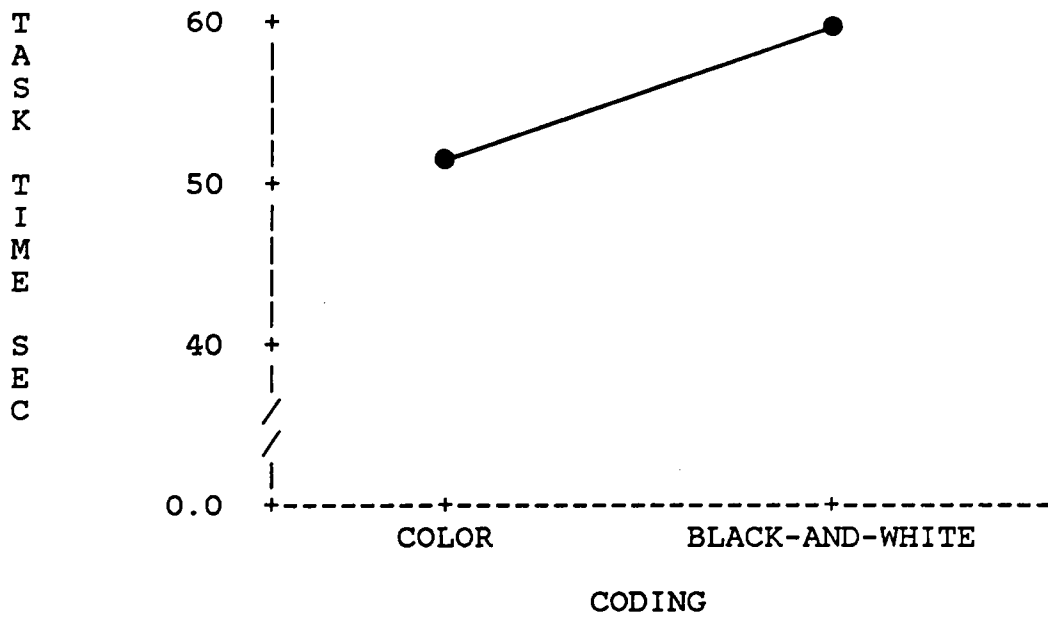


Figure 8. Effect of coding, point-comparison experiment, on task completion time.

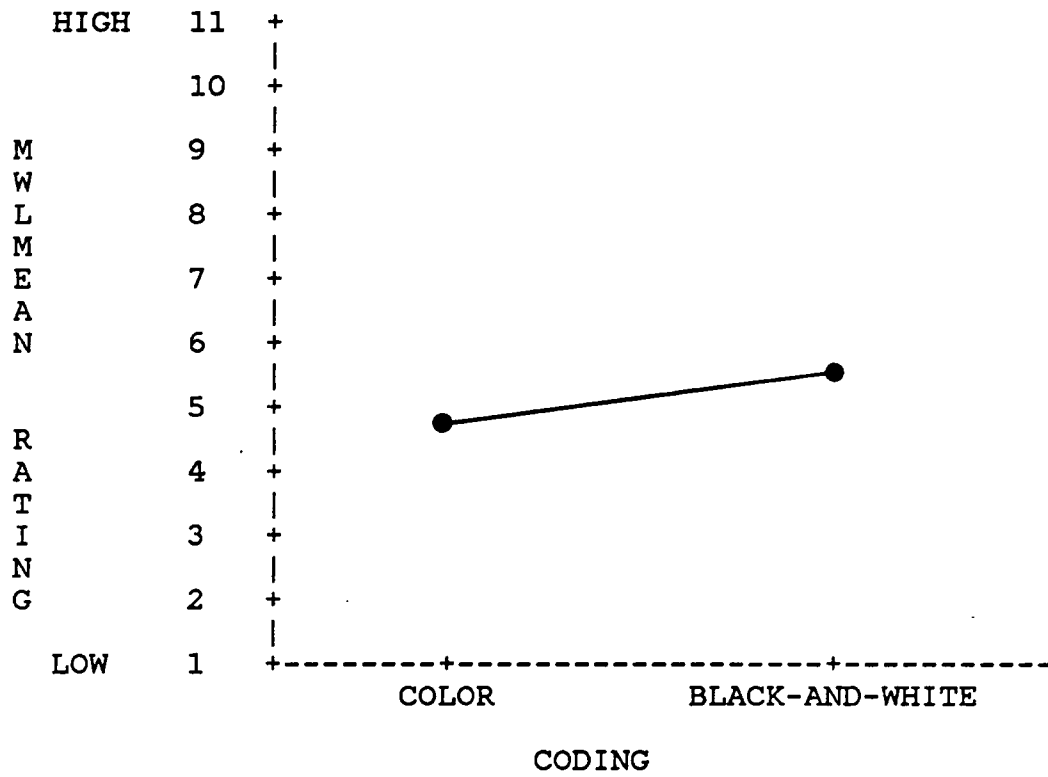


Figure 9. Effect of coding, point-comparison experiment, on MWLMEAN rating.

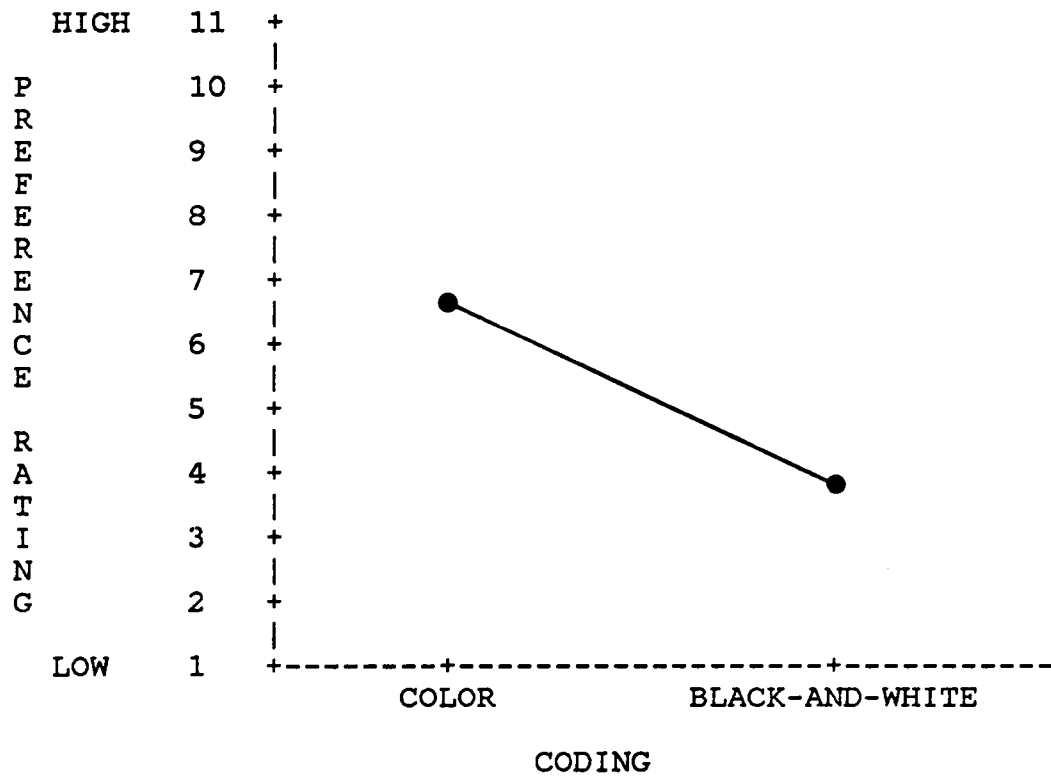


Figure 10. Effect of coding, point-comparison experiment, on preference rating.

Significance for the main effect of complexity was found for all dependent measures: percent error ( $F(1,28)=33.05, p=0.0001$ ), time ( $F(1,28)=172.74, p=0.0001$ ), MWLMEAN ( $F(1,28)=160.44, p=0.0001$ ), and preference ( $F(1,28)=105.50, p=0.0001$ ). As shown in Table 26, and Figures 11, 12, 13, and 14, the high complexity condition produced more errors, took longer to complete, had higher MWLMEAN ratings, and was, as expected, less preferred.

TABLE 26

Dependent Measure Mean Scores, Point-Comparison Experiment,  
for Levels of Complexity

---

	%ERROR	TIME	MWLMEAN	PREFERENCE
HIGH COMPLEXITY	25.78	73.96	6.33	4.32
LOW COMPLEXITY	7.03	36.56	3.89	6.46

---

where: TIME is in units of seconds

MWLMEAN is an interval scale in units of 1 (low) to 11 (high)

PREFERENCE is an interval scale in units of 1 (low) to 11 (high)

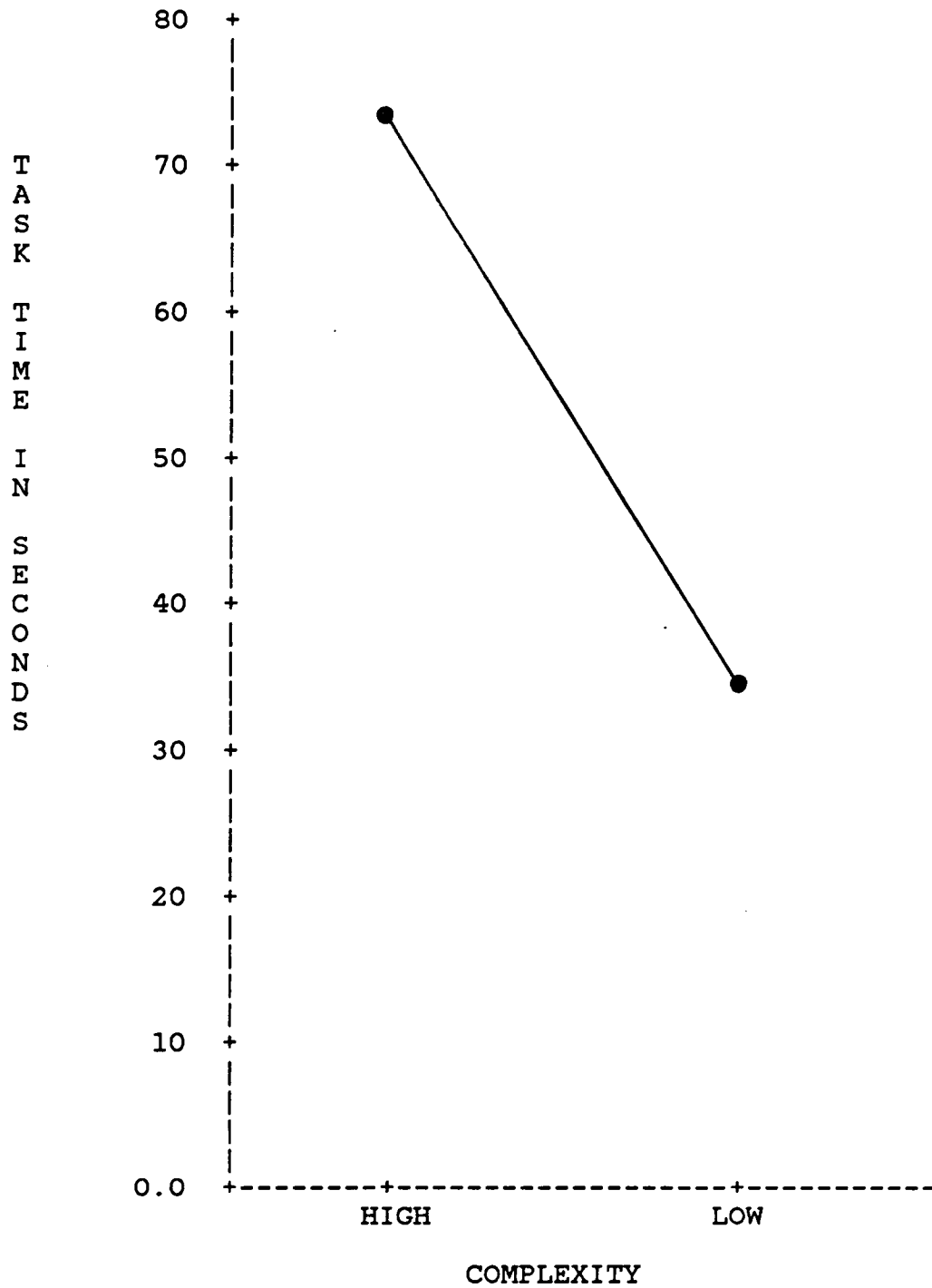


Figure 11. Effect of complexity, point-comparison experiment, on task completion time.



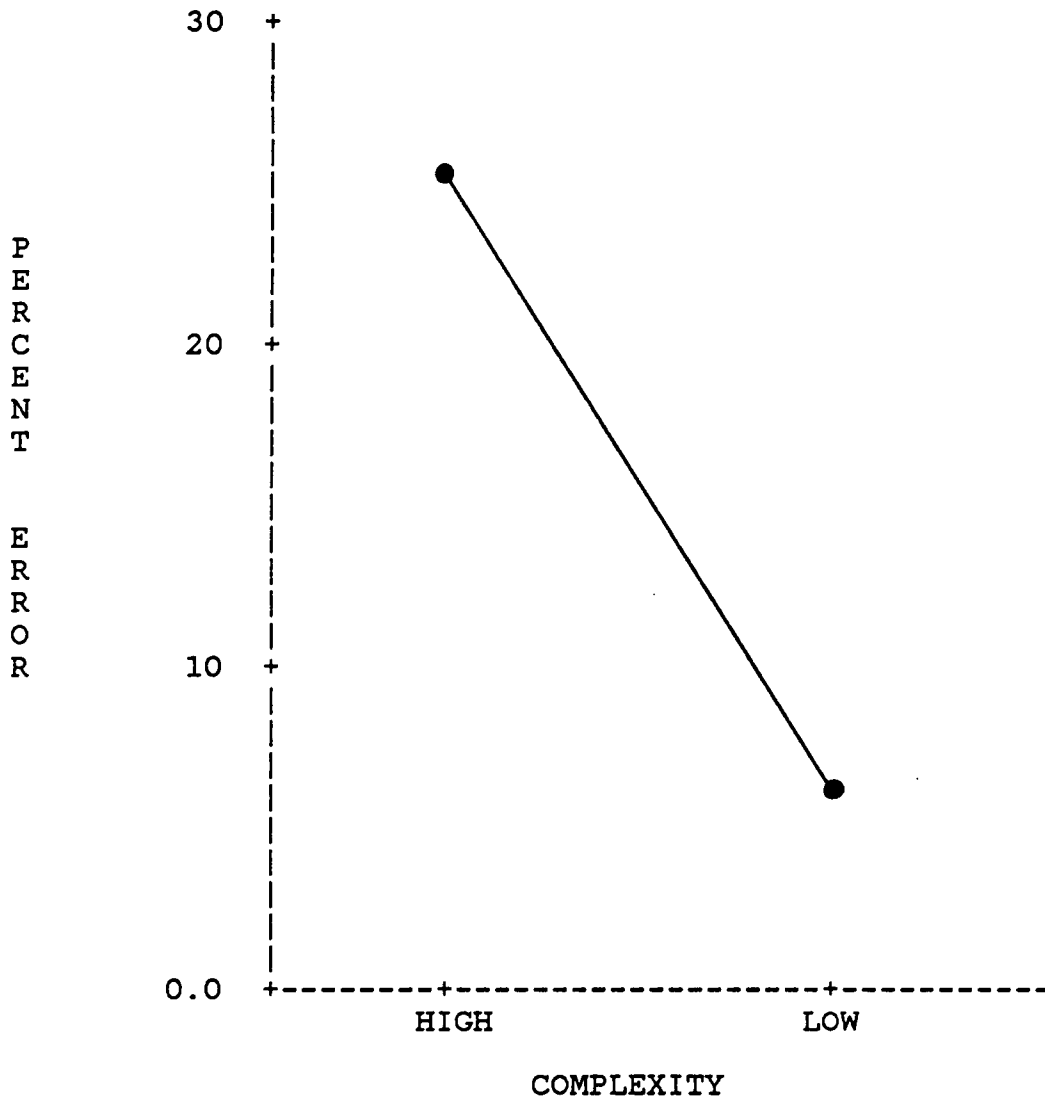


Figure 12. Effect of complexity, point-comparison experiment, on percent error score.

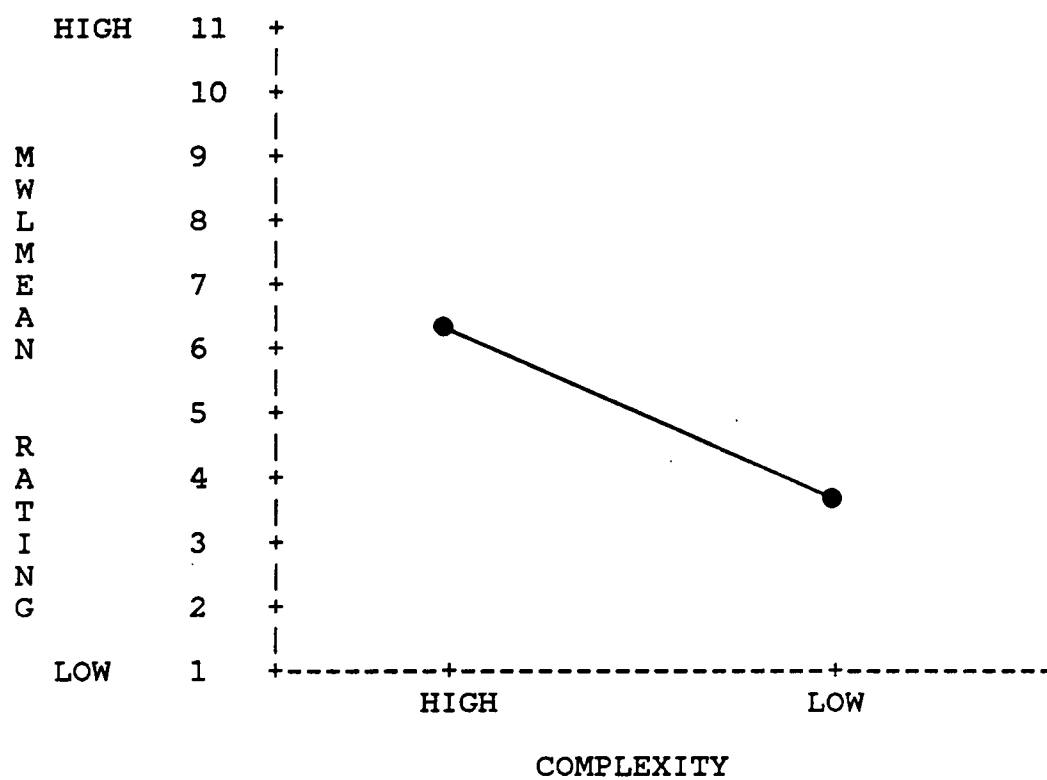


Figure 13. Effect of complexity, point-comparison experiment on MWLMEAN rating.

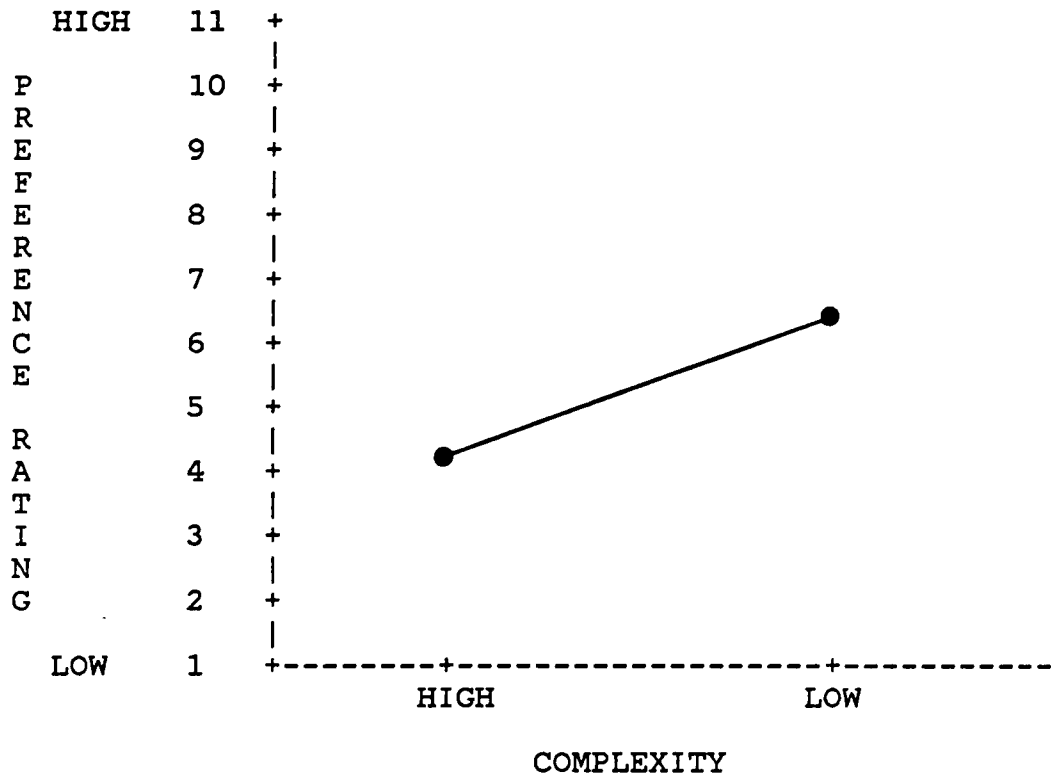


Figure 14. Effect of complexity, point-comparison experiment, on preference rating.

The graph-by-coding interaction was significant only for MWLMEAN ( $F(3,28)=3.26, p=0.0360$ ). The results of the Newman-Keuls analysis for these cell means (Table 27) indicated that black-and-white line and three-dimensional bar graphs had significantly higher MWLMEAN ratings than all other graphs, but were not significantly different from each other. In addition, color-coded point-plot graphs were rated significantly lower on this measure than black-and-white, or color-coded line and three-dimensional bar graphs. Finally, black-and-white point-plot graphs had significantly lower MWLMEAN ratings than color-coded line graphs. These results are displayed graphically in Figure 15.

TABLE 27

Newman-Keuls Analysis , Point-Comparison Experiment, of  
MWLMEAN for the Graph-by-Coding Interaction \*

GRAPH	PNT	PNT	BAR	3DB	BAR	LIN	LIN	3DB
CODING	COL	B&W	COL	COL	B&W	COL	B&W	B&W
MWL- MEAN	4.38	4.45	4.82	5.08	5.09	5.21	5.75	6.12
	_____							
		_____						
			_____					
						_____		

\* Treatments underlined by a common line do not differ from each other at  $p \leq 0.05$

where:      BAR = Bar graph  
              LIN = Line graph  
              PNT = Point-plot graph  
              3DB = Three-dimensional bar graph  
              B&W = Black-and-white coding  
              COL = Color-coding

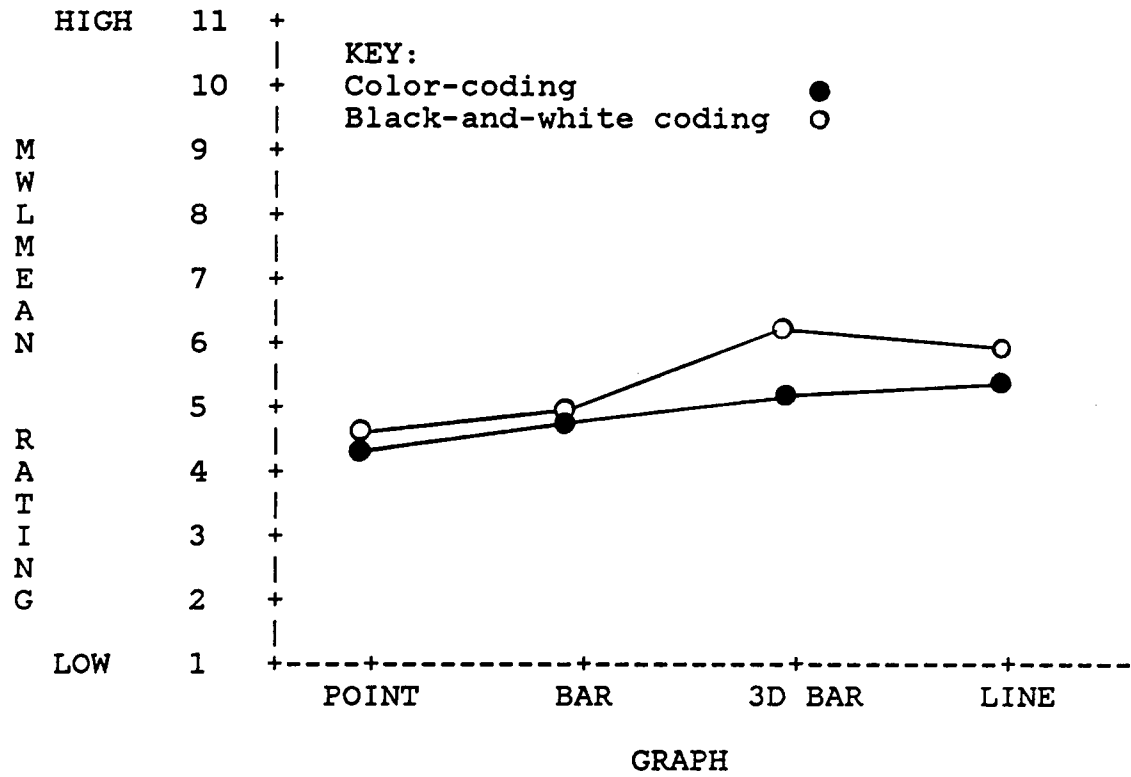


Figure 15. Effect of graph-type-by-coding interaction, point-comparison experiment, on MWLMEAN.

Trend-reading experiment. Summary tables for the four ANOVA analyses conducted for this experiment are shown in Tables 28, 29, 30, and 31. The results indicate that the main effect of coding was significant for the dependent measures of time ( $F(1,28)=7.35, p=0.0113$ ), MWLMEAN ( $F(1,28)=8.44, p=0.0071$ ), and preference ( $F(1,28)=59.32, p=0.0001$ ). Results are similar in nature to the two decision-tasks earlier discussed. Subjects took less time to complete the tasks when using color-coded graphs, rated them as requiring less overall mental effort (MWLMEAN), and preferred them over the black-and-white coded graphs (Table 32, and Figures 16, 17, and 18).

TABLE 28

ANOVA Summary Table, Trend-Reading Experiment, for the  
Dependent Measure Time

<u>Source</u>	<u>df</u>	<u>SS</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>				
Graph (G)	3	2573.385	1.22	0.3210
Subject (S)/G	28	19697.341		
<u>Within-Subject</u>				
Coding (CO)	1	598.487	7.35	0.0113 *
G X CO	3	323.078	1.32	0.2865
CO X S/G	28	2278.638		
Complexity (CM)	1	43551.762	190.80	0.0001 *
G X CM	3	787.126	1.15	0.3464
CM X S/G	28	6391.148		
CO X CM	1	97.338	0.93	0.3437
G X CO X CM	3	309.798	0.98	0.4144
CO X CM X S/G	28	2938.389		
<u>Total</u>	<u>127</u>			

\* significant at  $\alpha = 0.05$



TABLE 29

ANOVA Summary Table, Trend-Reading Experiment, for the  
Dependent Measure Percent Error

<u>Source</u>	<u>df</u>	<u>SS</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>				
Graph (G)	3	0.063	2.00	0.1372
Subject (S)/G	28	0.298		
<u>Within-Subject</u>				
Coding (CO)	1	0.004	0.68	0.4174
G X CO	3	0.079	4.09	0.0158 *
CO X S/G	28	0.181		
Complexity (CM)	1	0.109	10.57	0.0030 *
G X CM	3	0.042	1.68	0.1948
CM X S/G	28	0.291		
CO X CM	1	0.004	0.71	0.4073
G X CO X CM	3	0.087	4.69	0.0089 **
CO X CM X S/G	28	0.173		
<u>Total</u>	<u>127</u>			

\* significant at  $\alpha = 0.05$

\*\* not significant in MANOVA, therefore not considered here

TABLE 30

ANOVA Summary Table, Trend-Reading Experiment, for the  
Dependent Measure MWLMEAN

<u>Source</u>	<u>df</u>	<u>SS</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>				
Graph (G)	3	11.224	1.47	0.2448
Subject (S)/G	28	71.404		
<u>Within-Subject</u>				
Coding (CO)	1	6.412	8.44	0.0071 *
G X CO	3	4.266	1.87	0.1573
CO X S/G	28	21.274		
Complexity (CM)	1	192.447	242.70	0.0001 *
G X CM	3	8.136	3.42	0.0308 **
CM X S/G	28	22.202		
CO X CM	1	0.168	0.94	0.3406
G X CO X CM	3	0.631	1.17	0.3385
CO X CM X S/G	28	5.031		
<u>Total</u>	<u>127</u>			

\* significant at  $\alpha = 0.05$

\*\* not significant in MANOVA, therefore not considered here

TABLE 31

ANOVA Summary Table, Trend-Reading Experiment, for the  
Dependent Measure Preference

<u>Source</u>	<u>df</u>	<u>SS</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>				
Graph (G)	3	40.278	1.30	0.2935
Subject (S)/G	28	288.876		
<u>Within-Subject</u>				
Coding (CO)	1	190.734	59.32	0.0001 *
G X CO	3	13.528	1.40	0.2628
CO X S/G	28	90.033		
Complexity (CM)	1	147.168	121.05	0.0001 *
G X CM	3	12.587	3.45	0.0298 **
CM X S/G	28	34.041		
CO X CM	1	0.039	0.05	0.8306
CO X CM X G	3	5.872	2.31	0.0983
CO X CM X S/G	28	23.759		
<u>Total</u>	<u>127</u>			

\* significant at  $\alpha = 0.05$

\*\* not significant in MANOVA, therefore not considered here

TABLE 32

Dependent Measure Mean Scores, Trend-Reading Experiment,  
for Levels of Coding

---

	TIME	MWLMEAN	PREFERENCE
BLACK & WHITE CODING	43.61	4.05	5.31
COLOR CODING	39.29	3.60	7.75

---

where: TIME is in units of seconds

MWLMEAN is an interval scale in units of 1 (low)  
to 11 (high)

PREFERENCE is an interval scale in units of 1 (low)  
to 11 (high)

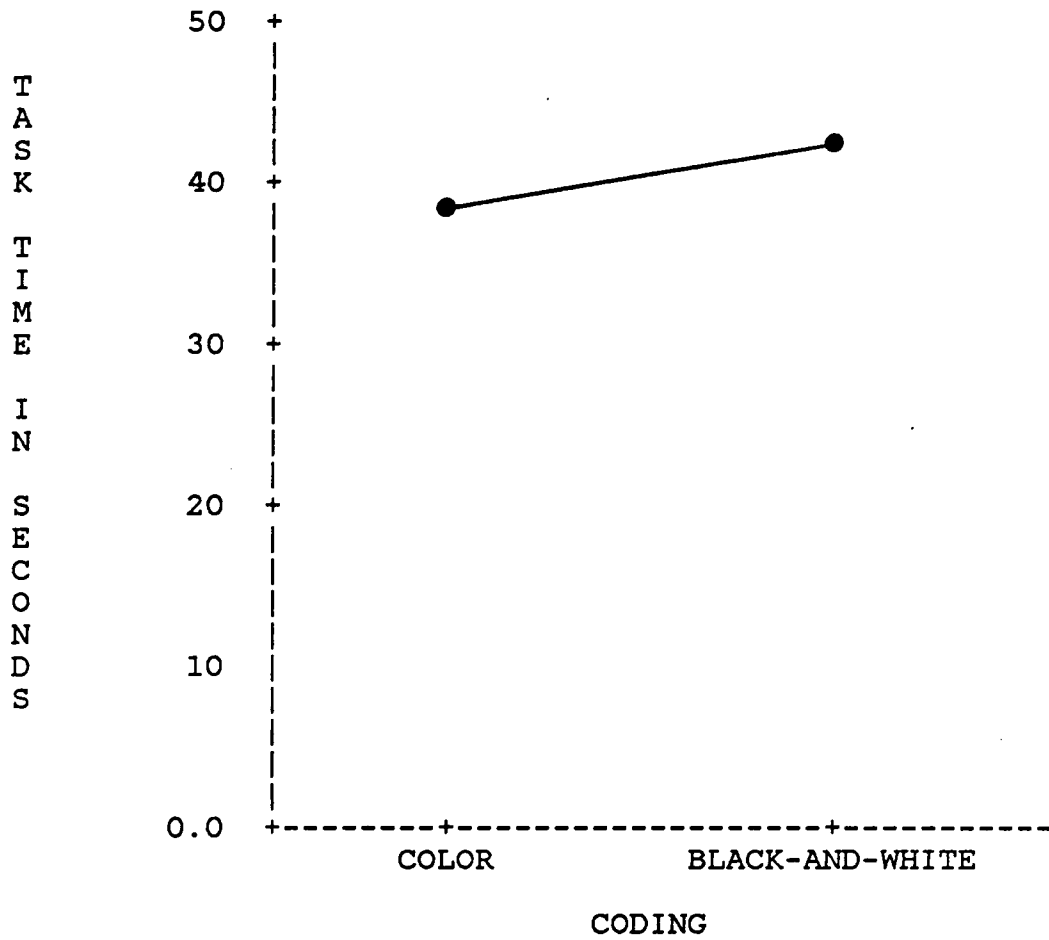


Figure 16. Effect of coding, trend-reading experiment, on task completion time.

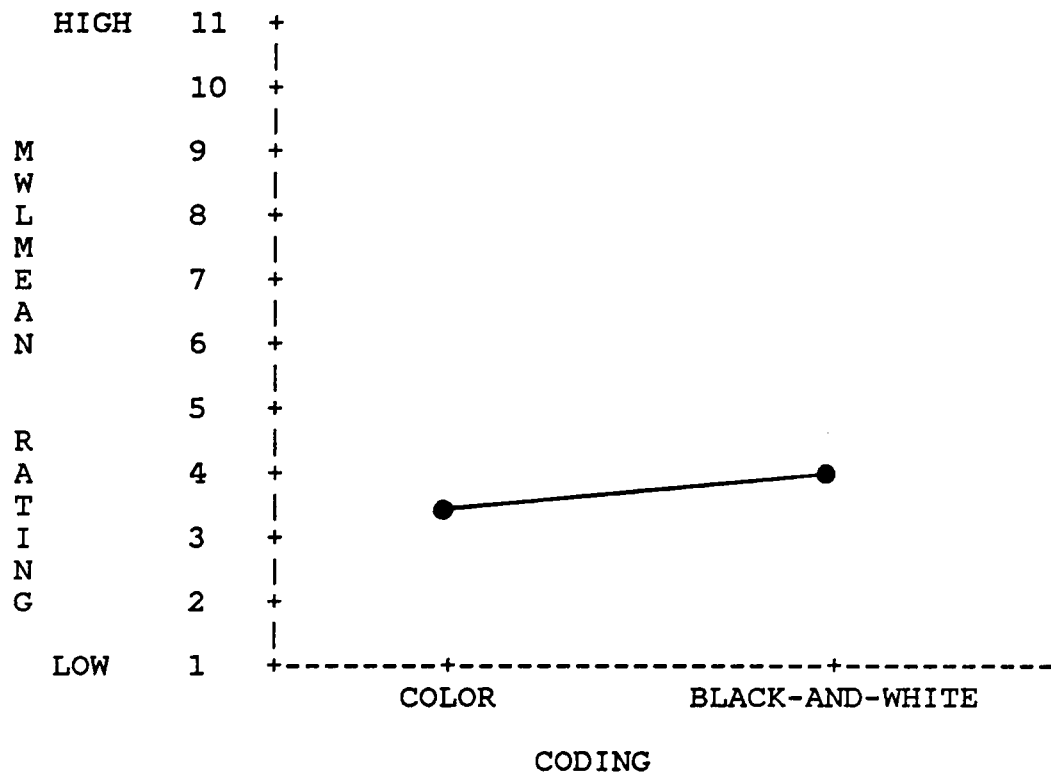


Figure 17. Effect of coding, trend-reading experiment, on MWLMEAN rating.

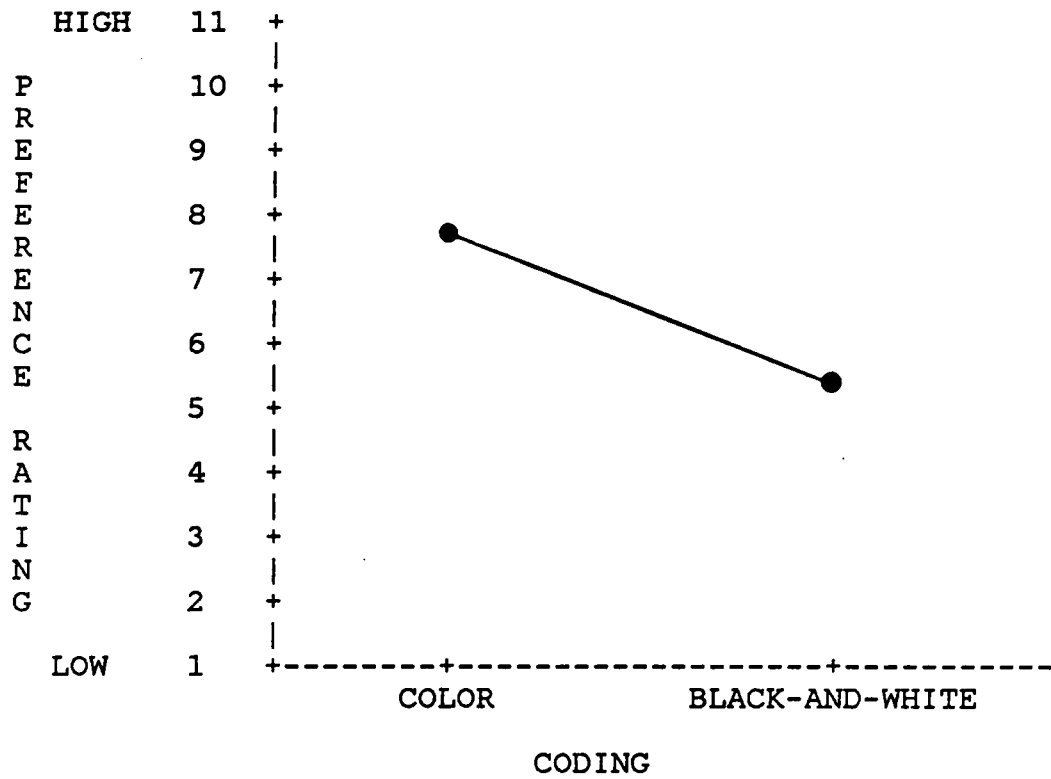


Figure 18. Effect of coding, trend-reading experiment, on preference rating.

Significance for the main effect of complexity was found for time ( $F(1,28)=190.80, p=0.0001$ ), MWLMEAN ( $F(1,28)=242.70, p=0.0001$ ), preference ( $F(1,28)=121.05, p=0.0001$ ), and percent errors ( $F(1,28)=10.57, p=0.0030$ ). The means (Table 33, and Figures 19, 20, 21, and 22) reflect greater errors, time in performing the task, and MWLMEAN scores for the high complexity condition. Meanwhile, low complexity graphs, as would be expected, were preferred over high complexity graphs.



TABLE 33

Dependent Measure Mean Scores, Trend-Reading Experiment,  
for Levels of Complexity

---

	%ERROR	TIME	MWLMEAN	PREFERENCE
HIGH COMPLEXITY	7.03	59.90	5.05	5.46
LOW COMPLEXITY	1.17	23.00	2.60	7.60

---

where: TIME is in units of seconds

MWLMEAN is an interval scale in units of 1 (low)  
to 11 (high)

PREFERENCE is an interval scale in units of 1 (low)  
to 11 (high)

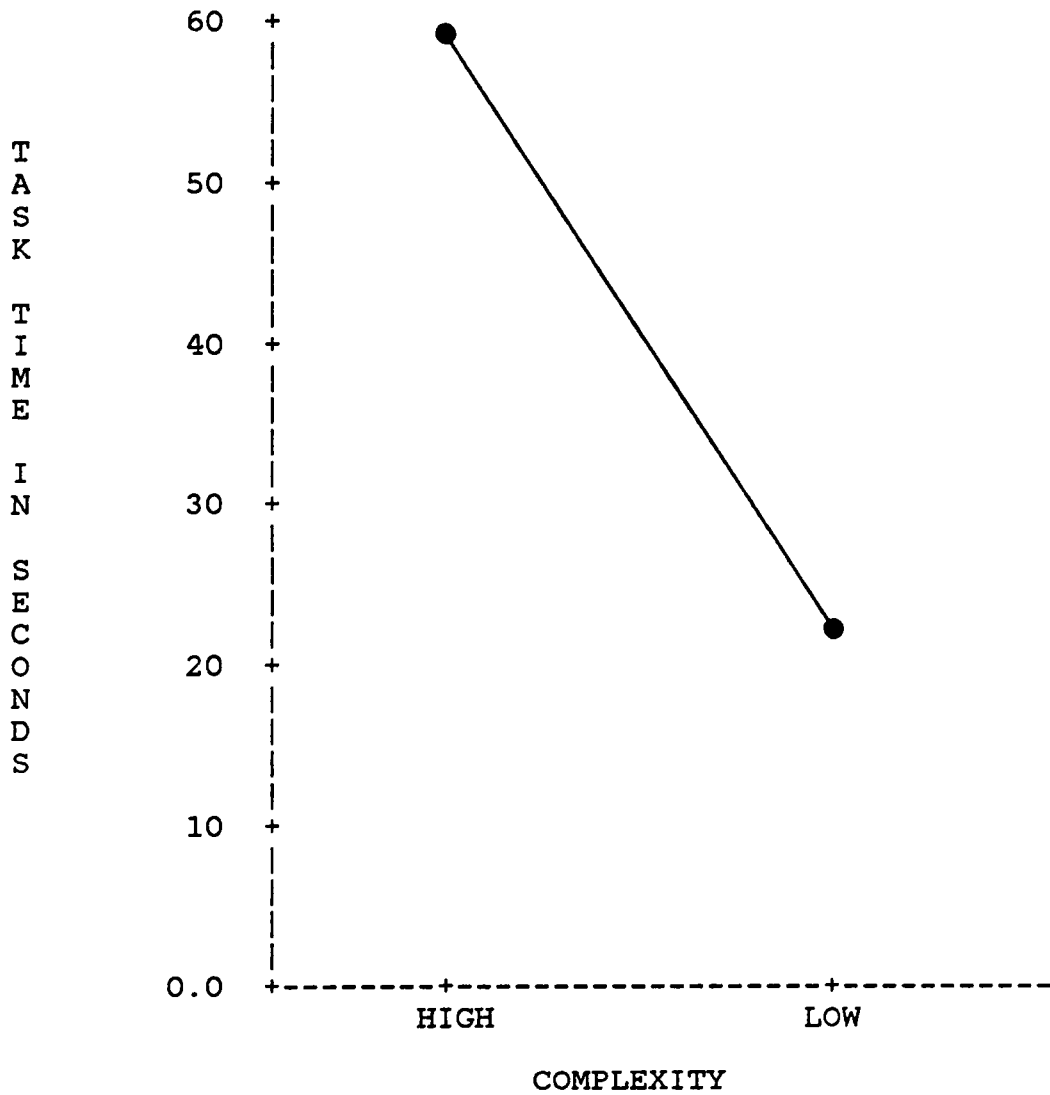


Figure 19. Effect of complexity, trend-reading experiment, on task completion time.

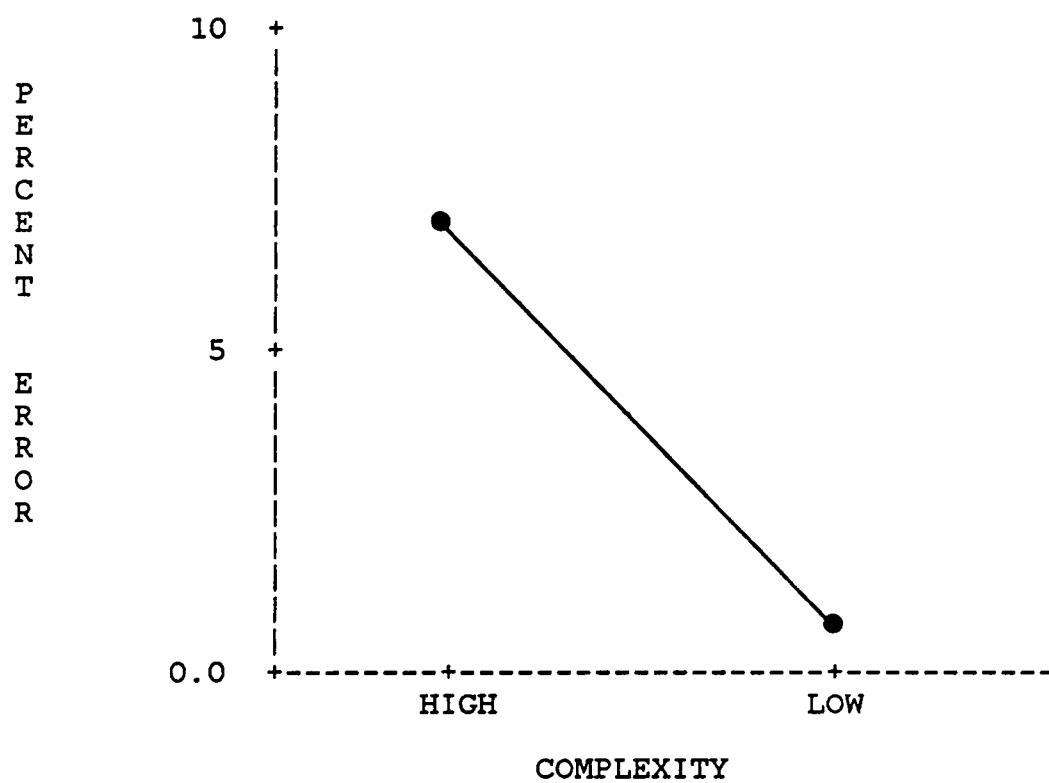


Figure 20. Effect of complexity, trend-reading experiment, on percent error score.

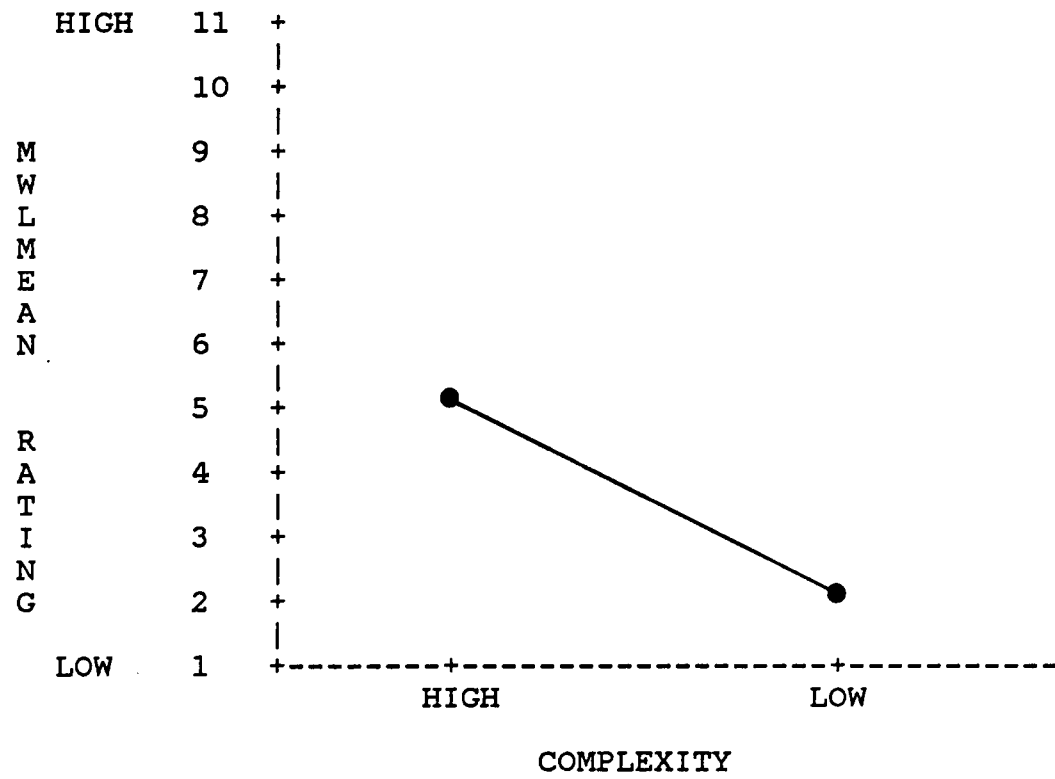


Figure 21. Effect of complexity, trend-reading experiment, on MWLMEAN rating.

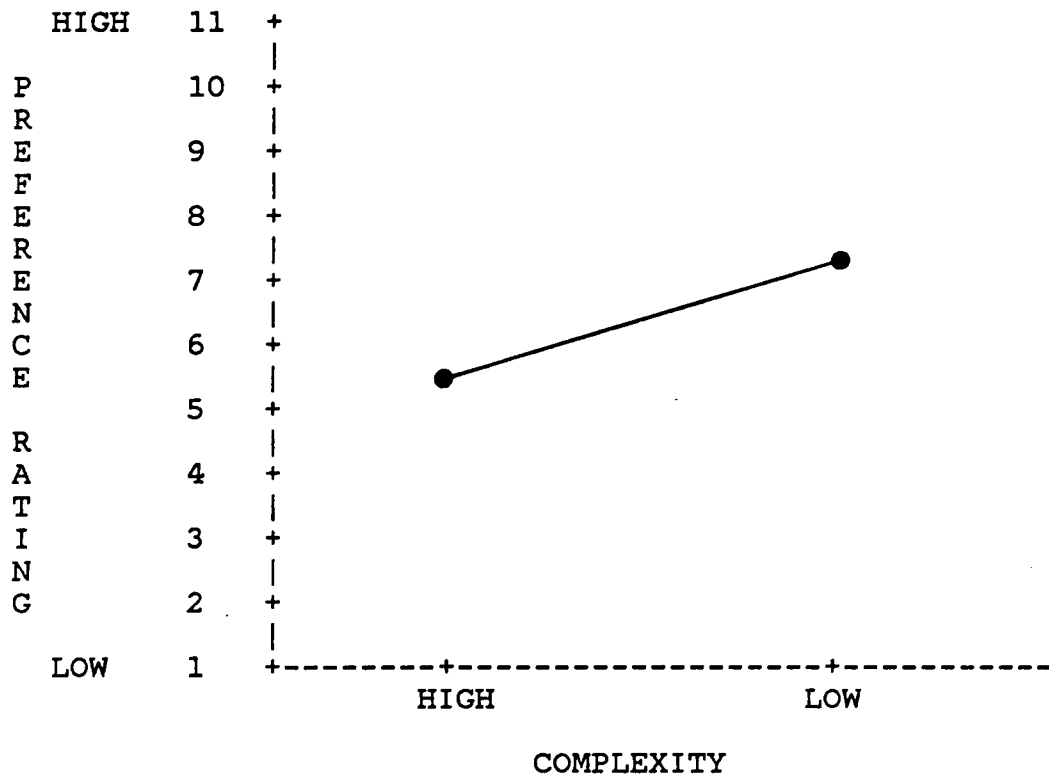


Figure 22. Effect of complexity, trend-reading experiment, on preference rating.

Investigation of the graph-by-coding interaction indicated significance only for the percent error measure ( $F(3,28)=4.09$ ,  $p=0.0158$ ). Results of a Newman-Keuls analysis of the treatment means are shown in Table 34, and indicate that color-coded three-dimensional bar graphs had a significantly greater amount of errors than color-coded line, bar, and point-plot graphs. Graphical representation of these results are shown in Figure 23.

TABLE 34

Newman-Keuls Analysis, Trend-Reading Experiment, of  
Percent Error Score for the Graph-by-Coding Interaction \*

---

GRAPH	BAR	PNT	LIN	LIN	PNT	3DB	BAR	3DB
CODING	COL	COL	COL	B&W	B&W	B&W	B&W	COL
MEAN	0.00	1.56	1.56	3.12	3.12	4.68	7.81	10.93
% ERROR								

---

\* Treatments underlined by a common line do not differ from each other at  $p \leq 0.05$

where:      BAR = Bar graph  
              LIN = Line graph  
              PNT = Point-plot graph  
              3DB = Three-dimensional bar graph  
              B&W = Black-and-white coding  
              COL = Color-coding

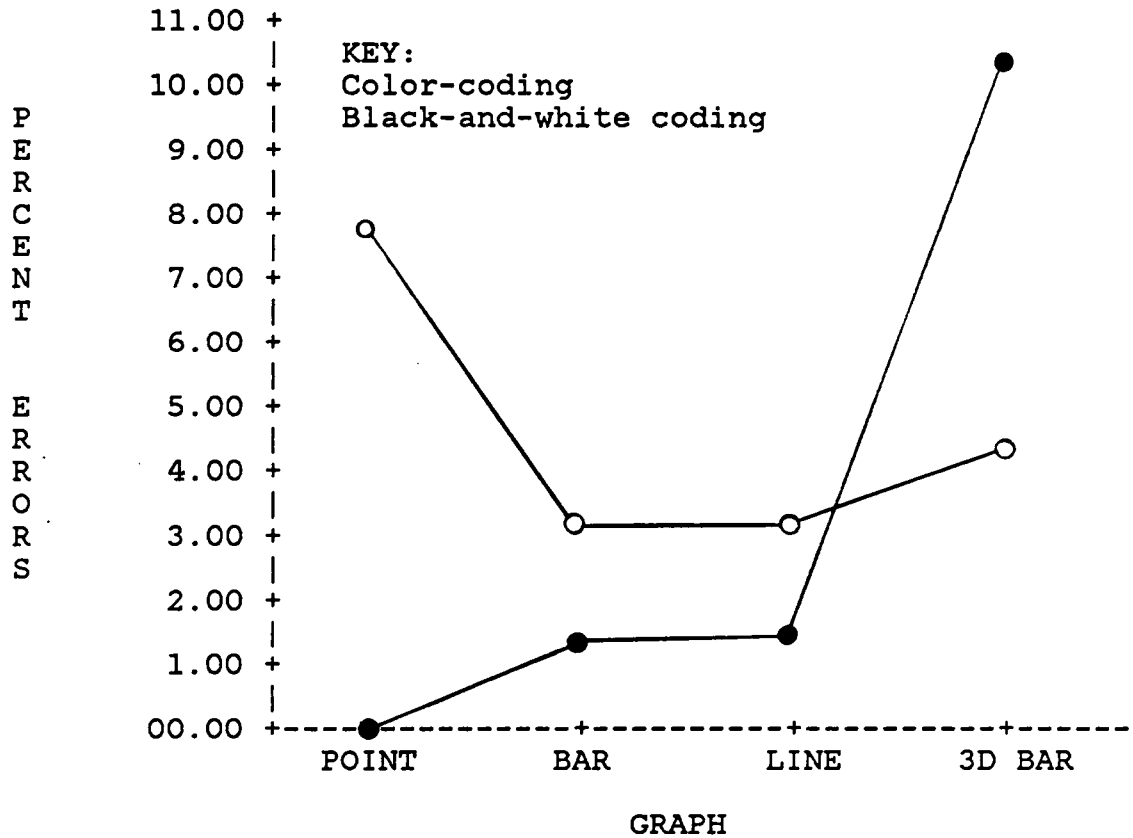


Figure 23. Effect of graph-type-by-coding interaction, trend-reading experiment, on percent error score.



Trend-comparison experiment. ANOVA results for this task (Tables 35, 36, 37, and 38) indicate significance of the main effect of coding for the dependent measures of time ( $F(1,28)=7.61$ ,  $p=0.0101$ ), MWLMEAN ( $F(1,28)=31.52$ ,  $p=0.0001$ ), and preference ( $F(1,28)=5.13$ ,  $p=0.0001$ ). Again, tasks using color-coded graphs required less time, were rated lower on MWLMEAN, and were preferred over black-and-white coding of graphs (see Table 39, and Figures 24, 25, and 26).

TABLE 35

ANOVA Summary Table, Trend-Comparison Experiment, for the  
Dependent Measure Time

<u>Source</u>	<u>df</u>	<u>SS</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>				
Graph (G)	3	2401.542	0.40	0.7521
Subject (S)/G	28	55650.021		
<u>Within-Subject</u>				
Coding (CO)	1	1328.868	7.61	0.0101 *
G X CO	3	36.519	0.90	0.4542
CO X S/G	28	4892.362		
Complexity (CM)	1	68970.837	128.24	0.0001 *
G X CM	3	144.435	0.09	0.9652
CM X S/G	28	15059.230		
CO X CM	1	36.519	0.17	0.6803
G X CO X CM	3	146.838	0.23	0.8730
CO X CM X S/G	28	5896.942		
<u>Total</u>	<u>127</u>			

\* significant at  $\alpha = 0.05$

TABLE 36

ANOVA Summary Table, Trend-Comparison Experiment, for the  
Dependent Measure Percent Error

<u>Source</u>	<u>df</u>	<u>SS</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>				
Graph (G)	3	0.438	2.04	0.1307
Subject (S)/G	28	2.005		
<u>Within-Subject</u>				
Coding (CO)	1	0.004	0.12	0.7369
G X CO	3	0.099	0.87	0.4703
CO X S/G	28	1.068		
Complexity (CM)	1	3.203	51.55	0.0001 *
G X CM	3	0.165	0.89	0.4595
CM X S/G	28	1.740		
CO X CM	1	0.012	0.38	0.5419
CO X CM X G	3	0.013	0.14	0.9369
CO X CM X S/G	28	0.896		
<u>Total</u>	<u>127</u>			

\* significant at  $\alpha = 0.05$

TABLE 37

ANOVA Summary Table, Trend-Comparison Experiment, for the  
Dependent Measure MWLMEAN

<u>Source</u>	<u>df</u>	<u>SS</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>				
Graph (G)	3	6.702	0.48	0.6957
Subject (S)/G	28	129.105		
<u>Within-Subject</u>				
Coding (CO)	1	8.846	31.52	0.0001 *
G X CO	3	1.103	1.31	0.2917
CO X S/G	28	7.857		
Complexity (CM)	1	136.228	188.22	0.0001 *
G X CM	3	8.030	3.70	0.0233 **
CM X S/G	28	20.265		
CO X CM	1	0.168	1.10	0.3031
G X CO X CM	3	0.729	1.59	0.2152
CO X CM X S/G	28	4.298		
<u>Total</u>	<u>127</u>			

\* significant at  $\alpha = 0.05$

\*\* not significant in MANOVA, therefore not considered here

TABLE 38

ANOVA Summary Table, Trend-Comparison Experiment, for the  
Dependent Measure Preference

<u>Source</u>	<u>df</u>	<u>SS</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>				
Graph (G)	3	49.570	2.42	0.0866
Subject (S)/G	28	190.804		
<u>Within-Subject</u>				
Coding (CO)	1	159.757	5.13	0.0001 *
G X CO	3	14.859	1.17	0.1878
CO X S/G	28	81.132		
Complexity (CM)	1	145.564	101.52	0.0001 *
G X CM	3	9.287	2.16	0.1152
CM X S/G	28	40.148		
CO X CM	1	1.642	2.16	0.1528
G X CO X CM	3	0.693	0.30	0.8223
CO X CM X S/G	28	21.289		
<u>Total</u>	<u>127</u>			

\* significant at  $\alpha = 0.05$

TABLE 39

Dependent Measure Mean Scores, Trend-Comparison  
Experiment, for Levels of Coding

---

	TIME	MWLMEAN	PREFERENCE
BLACK & WHITE CODING	81.24	6.31	3.82
COLOR CODING	74.79	5.79	6.05

---

where: TIME is in units of seconds

MWLMEAN is an interval scale in units of 1 (low)  
to 11 (high)

PREFERENCE is an interval scale in units of 1 (low)  
to 11 (high)

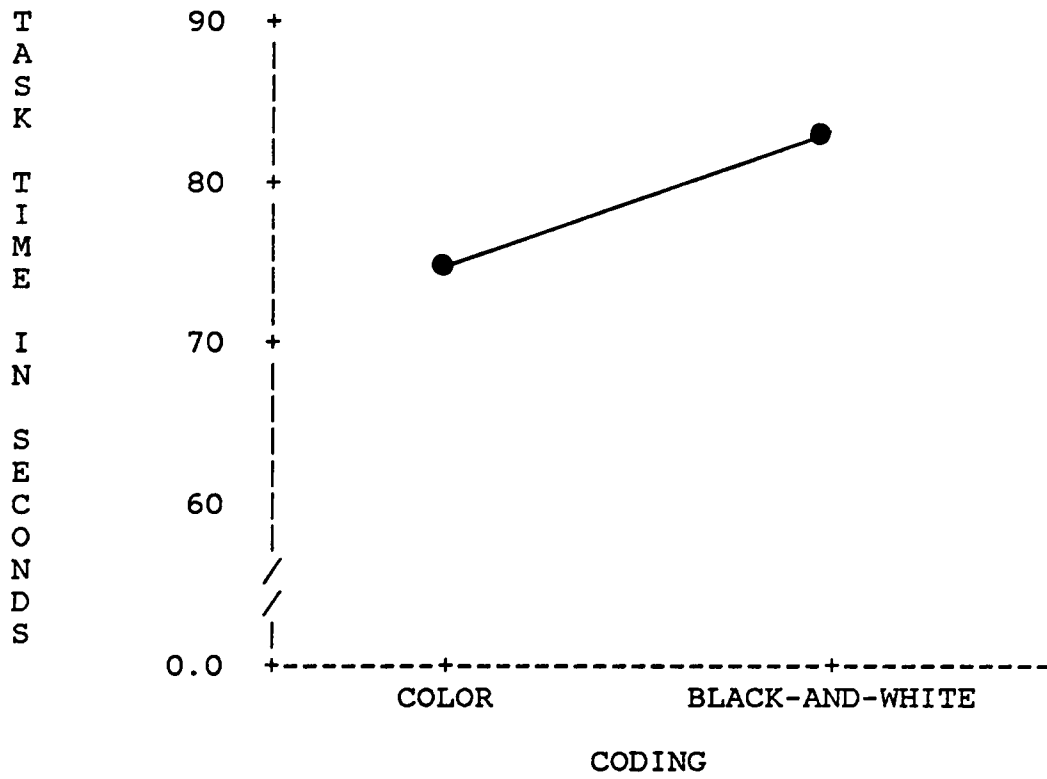


Figure 24. Effect of coding, trend-comparison experiment, on task completion time.

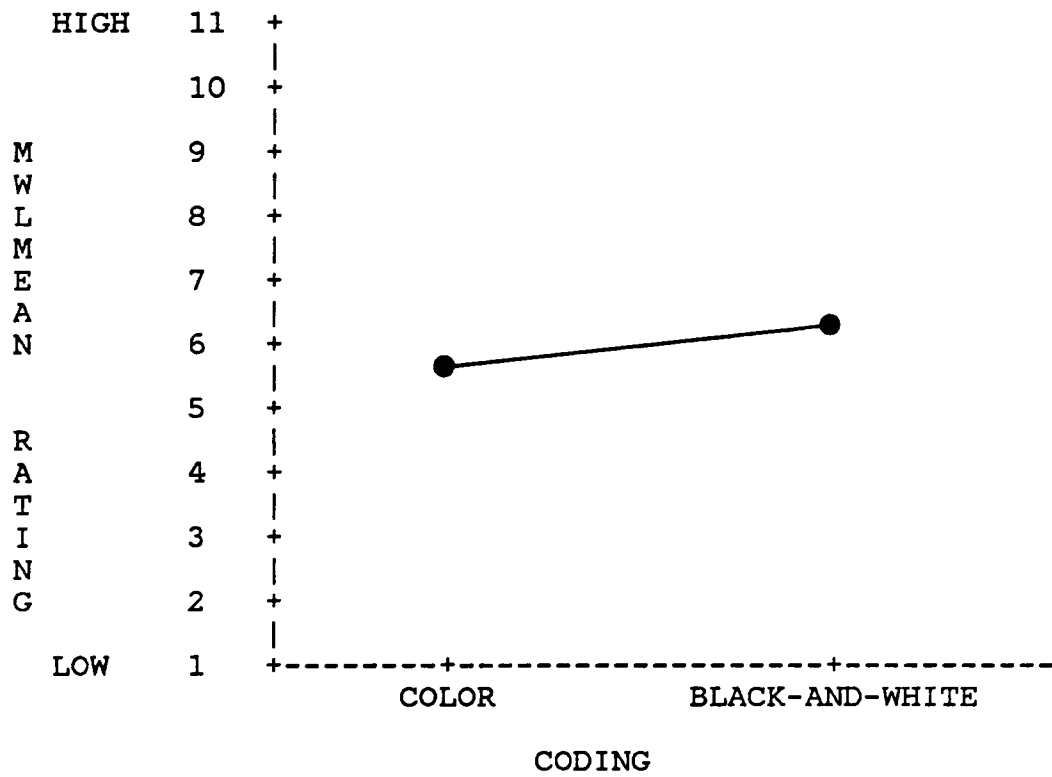


Figure 25. Effect of coding, trend-comparison experiment, on MWLMEAN rating.



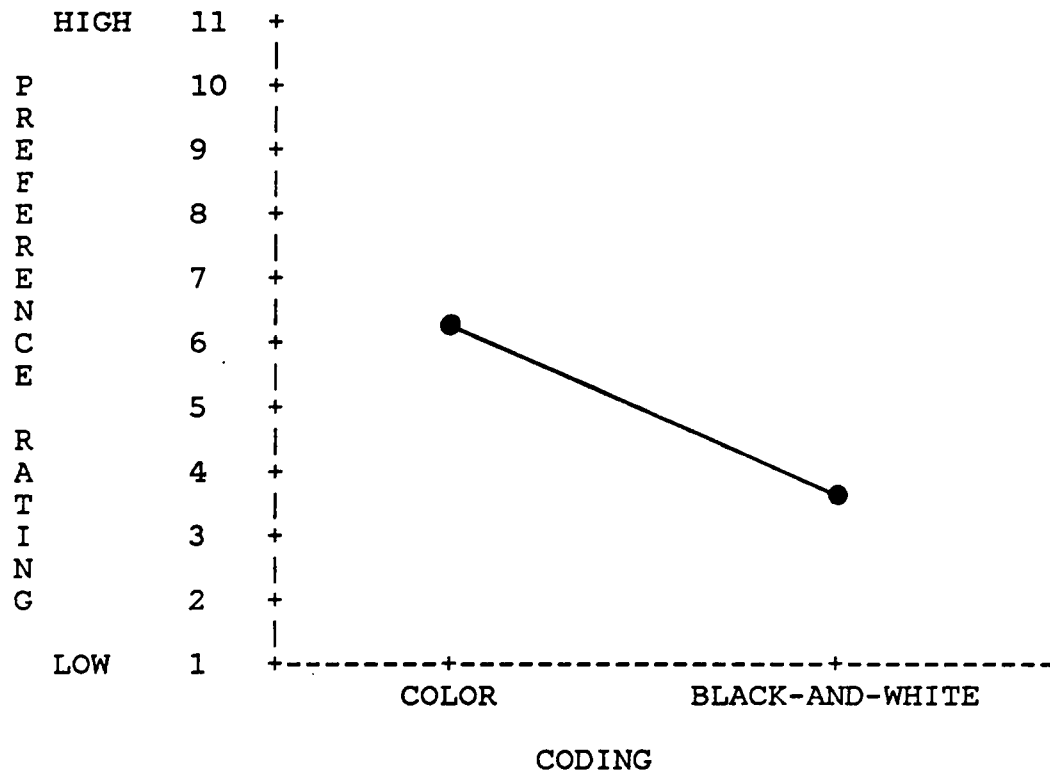


Figure 26. Effect of coding, trend-comparison experiment, on preference rating.

The main effect of complexity was significant for all dependent measures, percent error ( $F(1,28)=51.55$ ,  $p=0.0001$ ), time ( $F(1,28)=128.24$ ,  $p=0.0001$ ), MWLMEAN ( $F(1,28)=188.22$ ,  $p=0.0001$ ), and preference ( $F(1,28)=101.52$ ,  $p=0.0001$ ). High complexity conditions had greater percent error scores, took more time to complete, had higher MWLMEAN ratings, and were less preferred (Table 40, and Figures 27, 28, 29, 30).

TABLE 40

Dependent Measure Mean Scores, Trend-Comparison Experiment,  
for Levels of Complexity

---

	%ERROR	TIME	MWLMEAN	PREFERENCE
HIGH COMPLEXITY	39.06	101.23	7.08	3.87
LOW COMPLEXITY	7.42	54.81	5.02	6.00

---

where: TIME is in units of seconds

MWLMEAN is an interval scale in units of 1 (low)  
to 11 (high)

PREFERENCE is an interval scale in units of 1 (low)  
to 11 (high)

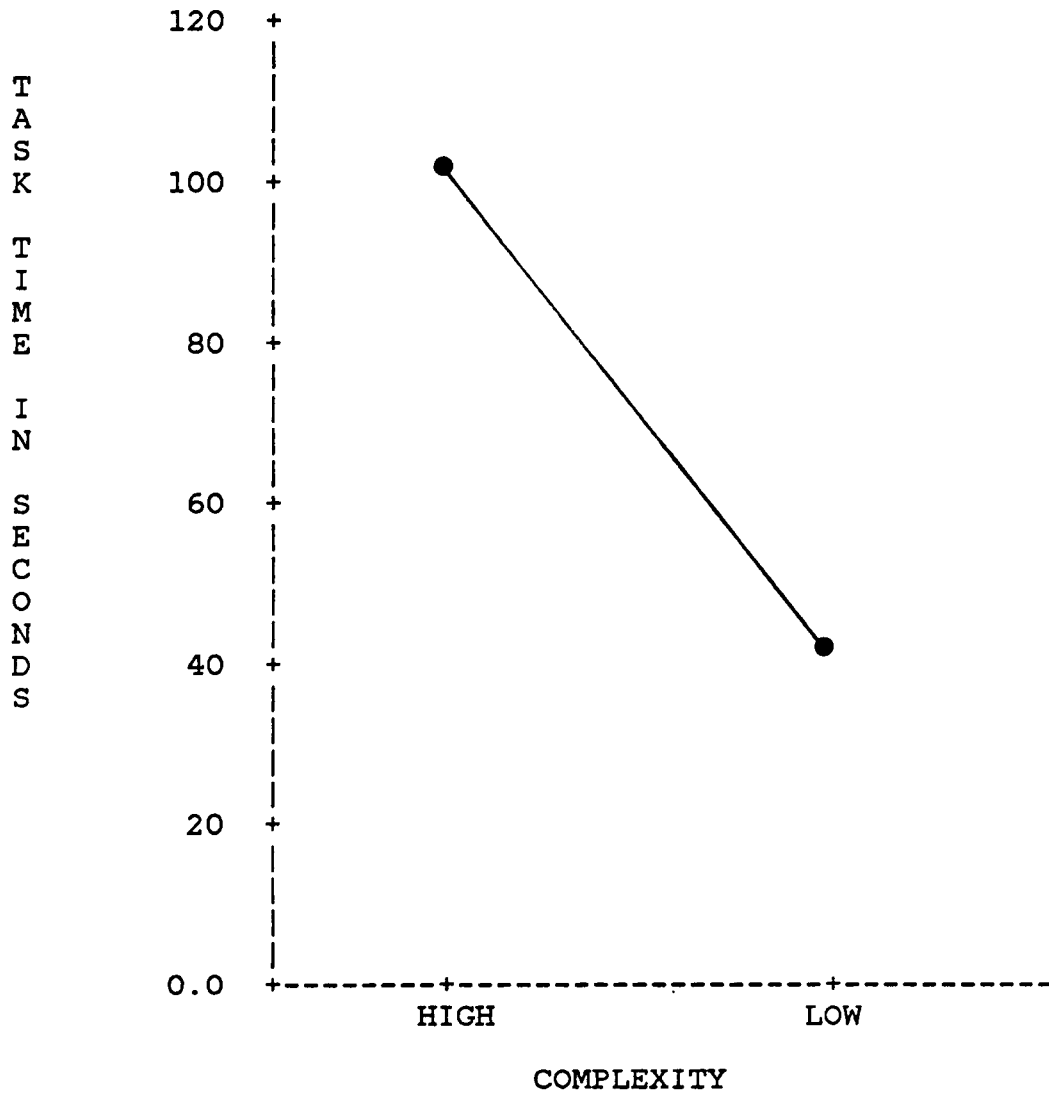


Figure 27. Effect of complexity, trend-comparison experiment, on task completion time.

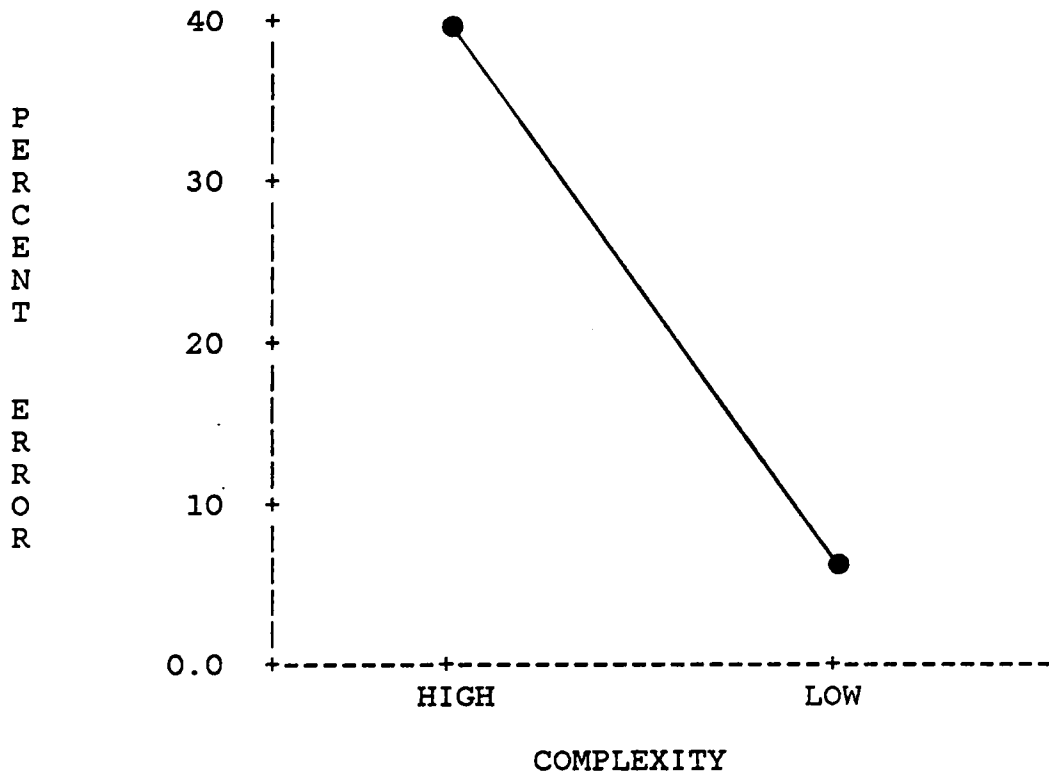


Figure 28. Effect of complexity, trend-comparison experiment, on percent error score.

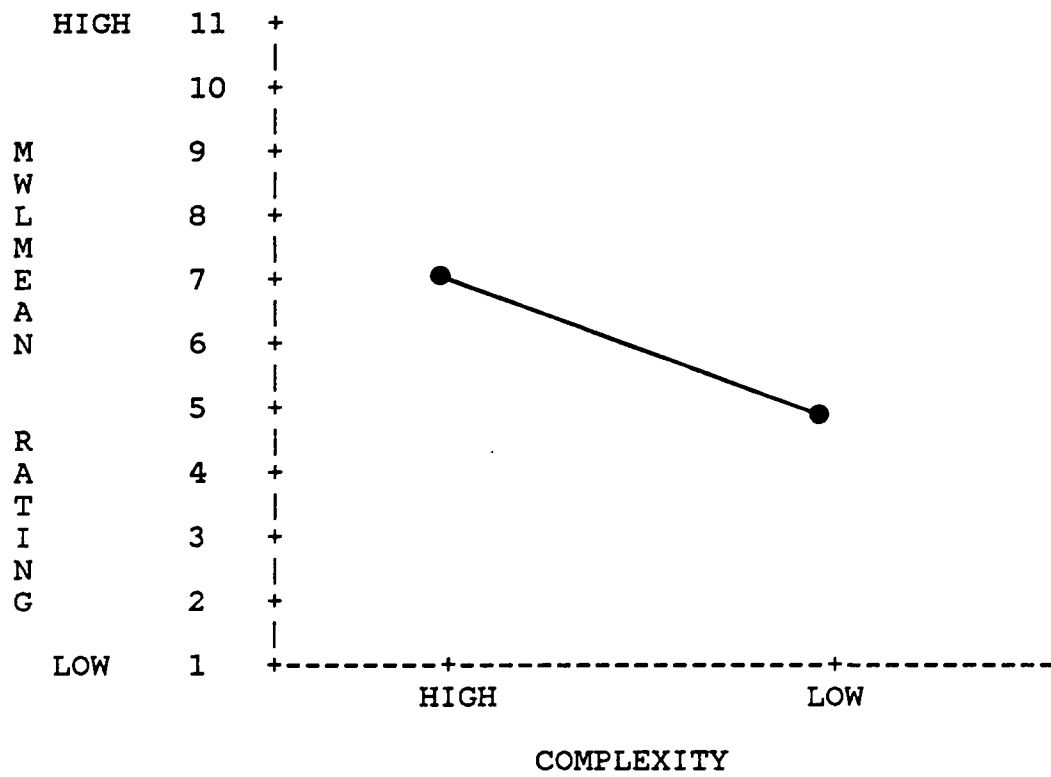


Figure 29. Effect of complexity, trend-comparison experiment, on MWLMEAN rating.

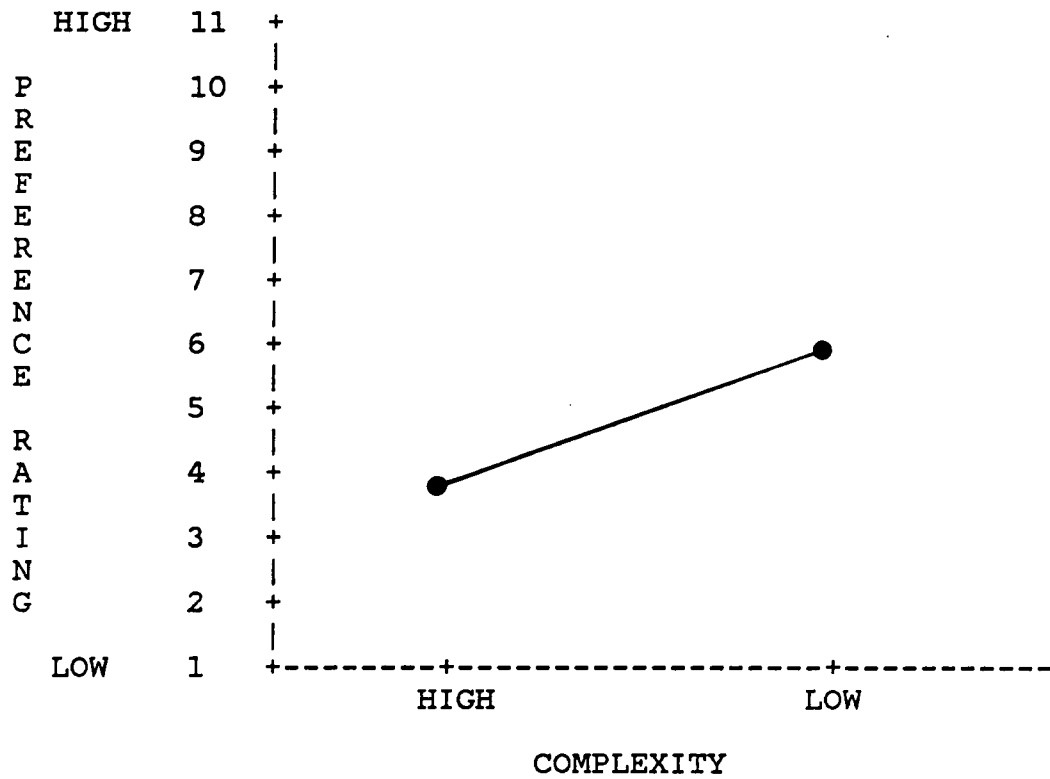


Figure 30. Effect of complexity, trend-comparison experiment, on preference rating.

It should be reiterated that there were several effects for which the ANOVA of a dependent measure showed significance, but the overall MANOVA did not, and these were not further investigated. As was indicated, this was done to control for excessive alpha error. Although it is important to maintain an acceptable experiment-wise protection level by limiting the consideration of significant effects beyond those in the MANOVA domain, important information about the data may be missed if too stringent a criterion is adopted. To maintain statistical rigor, but still investigate the data further, a different MANOVA strategy was adopted in a second set of analyses. These are discussed in the next section.

#### MANOVA Analysis of Objective Versus Subjective Measures

In contrast to lumping all dependent measures together in one large MANOVA as before, separate MANOVA's were also conducted on the objective and subjective dependent measures for each task. This breakdown was believed to be the most logical separation among measures, and also the one with the least chance of interdependence. Objective measures included time to complete the task and errors in performing the task, while subjective measures included the subjective mental workload mean score (MWLMEAN) and



preference. For the purposes of this study it was important to isolate which independent variables were responsible for differences in performance. It was these objective measures of performance that directly reflect upon the utility of a graphical technique by indicating probable savings in time and the reduction of errors and inaccuracies. These measures were included in one MANOVA because of the distinct possibility that time and error measures covaried. Secondly, past research has suggested that there might be differences between subjects' preference or subjective perceptions of task difficulty and the results of objective measures of performance in the task.

Due to the probable correlation of time and error scores it was important to look at ANOVA results only for effects that were significant for the MANOVA in which they were combined. That is, a subject who took the time to insure that his or her answers were accurate would be more likely to have fewer errors than a subject who rushed through each task. Thus, separate ANOVA analysis of these measures might be misleading, while MANOVA procedures are capable of accounting for this type of interdependence.

Grouping the subjective measures of MWLMEAN and preference together for a MANOVA analysis was done for

comparative purposes with the MANOVA for objective measures. In this way, it was possible to determine if there were any differences in the power of these two types of measures to detect the effects of independent variables. The subjective measures were constructed to measure two separate and discrete parameters that are not necessarily related to each other. The intended purpose of the MWLMEAN measures was to estimate composite mental workload by tapping several of its underlying dimensions. On the other hand, the preference rating scale was worded to measure the distinctly different quality of how much a subject liked or was satisfied with a graphical technique for use in a given task, perhaps just by looking at it, or using it. Thus these measures are not, by definition, linked together, as were the objective measures of time and error, and could be a source of separate and different types of information. If so, this effect could be demonstrated in the subsequent univariate analyses on each separate measure.

Point-reading experiment. MANOVA analysis of the objective measures (Table 41) for the point-reading task indicated significance only for the graph main effect ( $F(6,54)=3.27, p=0.0081$ ), however the graph-by-coding interaction was almost significant ( $F(6,54)=2.27, p=0.0501$ ). As discussed earlier, the location for this effect was restricted to the dependent measure of time to

complete the task ( $F(3,28)=4.79$ ,  $p=0.0081$ ), shown in Table 14.

The MANOVA of subjective measures (Table 42) detected significant results for coding ( $F(2,27)=30.00$ ,  $p=0.0001$ ), complexity ( $F(2,27)=10.26$ ,  $p=0.0005$ ), and the graph-by-coding interaction ( $F(6,54)$ ,  $p=0.0479$ ), as did the MANOVA done earlier on all four dependent variables. MWLMEAN (Table 16) was found significant for complexity only ( $F(1,28)=9.31$ ,  $p=0.0050$ ). Preference (Table 17) was significant for the main effect of coding ( $F(1,28)=61.11$ ,  $p=0.0001$ ), and complexity ( $F(1,28)=18.48$ ,  $p=0.0002$ ).

In summary, for the point-reading experiment, no additional statistically-significant results were found in the two separate MANOVA's of objective and subjective measures that were not present in the overall (objective plus subjective measures) MANOVA.

TABLE 41

MANOVA Summary Table, Point-Reading Experiment, for the Dependent Measures of RMS Error, Time

<u>Source</u>	<u>dv</u>	<u>df</u> H	<u>df</u> E	<u>U</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>						
Graph (G)	2	3	28	0.5379	3.27	0.0081 *
Subject (S)/G	(Error Term for G)					
<u>Within-Subject</u>						
Coding (CO)	2	1	28	0.9949	0.07	0.9336
G X CO	2	3	28	0.6376	2.27	0.0501
CO X S/G	(Error Term for CO, G X CO)					
Complexity (CM)	2	1	28	0.8611	2.18	0.1328
G X CM	2	3	28	0.8197	0.94	0.4742
CM X S/G	(Error Term for CM, G X CM)					
CO X CM	2	1	28	0.9889	0.15	0.8606
G X CO X CM	2	3	28	0.9239	0.36	0.8990
CO X CM X S/G	(Error Term for CO X CM, G X CO X CM)					

\* significant at  $\alpha = 0.05$

where: dv = number of dependent measures

$\frac{df}{H}$  = degrees of freedom for treatment effect

$\frac{df}{E}$  = degrees of freedom for error effect

U = Wilk's likelihood ratio statistic

$$\frac{|E|}{|E + H|}, \text{ where:}$$

|E| = determinant of sum of squares and cross-products for error

|E + H| = determinant of the sum of the sum of squares and cross-products matrix for error, and the sum of squares and cross-products matrix for treatment

TABLE 42

MANOVA Summary Table, Point-Reading Experiment, for  
Dependent Measures of MWLMEAN, Preference

<u>Source</u>	<u>dv</u>	<u>df</u> H	<u>df</u> E	<u>U</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>						
Graph (G)	2	3	28	0.6993	1.76	0.1246
Subject (S)/G	(Error Term for G)					
<u>Within-Subject</u>						
Coding (CO)	2	1	28	0.3103	30.00	0.0001 *
G X CO	2	3	28	0.6347	2.30	0.0479 *
CO X S/G	(Error Term for CO, G X CO)					
Complexity (CM)	2	1	28	0.5682	10.26	0.0005 *
G X CM	2	3	28	0.7206	1.60	0.1646
CM X S/G	(Error Term for CM, G X CM)					
CO X CM	2	1	28	0.8955	1.57	0.2257
G X CO X CM	2	3	28	0.9153	0.41	0.8712
CO X CM X S/G	(Error Term for CO X CM, G X CO X CM)					

\* significant at  $\alpha = 0.05$

where: dv = number of dependent measures

df<sub>H</sub> = degrees of freedom for treatment effect

df<sub>E</sub> = degrees of freedom for error effect

U = Wilk's likelihood ratio statistic

$$\frac{|E|}{|E + H|}, \text{ where:}$$

|E| = determinant of sum of squares and cross-products for error

|E + H| = determinant of the sum of the sum of squares and cross-products matrix for error, and the sum of squares and cross-products matrix for treatment

Point-comparison experiment. Results of the MANOVA for objective measures (Table 43) indicated significance for coding ( $F(2,27)=8.12$ ,  $p=0.0017$ ), and complexity ( $F(2,27)=88.86$ ,  $p=0.0001$ ).

The MANOVA analysis of subjective measures (Table 44) showed significance for complexity ( $F(2,27)=83.98$ ,  $p=0.0001$ ) and coding ( $F(2,27)=34.05$ ,  $p=0.0001$ ), graph-by-coding ( $F(6,54)=3.06$ ,  $p=0.0120$ ), and graph ( $F(6,54)=2.33$ ,  $p=0.0452$ ). The graph main effect was the only additional effect that was not previously found significant in the overall MANOVA discussed earlier. Further investigation of ANOVA's for the point-comparison experiment (see Tables 23, and 24) indicated significance of graph only for MWLMEAN ( $F(3,28)=3.12$ ,  $p=0.0417$ ). Somewhat surprisingly, a Newman-Keuls analysis (Table 45) revealed no significant differences between graph types. This is most likely due to the borderline case where the increased protection level of the Newman-Keuls test nullified differences found in the ANOVA. Further analysis of these means with the Duncan Multiple Range test (Table 46) indicated that point-plot graphs were rated significantly lower on MWLMEAN rating than line or three-dimensional bar graphs. This data is shown graphically in Figure 31.

TABLE 43

MANOVA Summary Table, Point-Comparison Experiment, for the Dependent Measures of Percent Error, Time

<u>Source</u>	<u>dy</u>	<u>df</u> H	<u>df</u> E	<u>U</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>						
Graph (G)	2	3	28	0.7269	1.56	0.1781
Subject (S)/G	(Error Term for G)					
<u>Within-Subject</u>						
Coding (CO)	2	1	28	0.6243	8.12	0.0017 *
G X CO	2	3	28	0.7582	1.34	0.2576
CO X S/G	(Error Term for CO, G X CO)					
Complexity (CM)	2	1	28	0.1318	88.86	0.0001 *
G X CM	2	3	28	0.8175	0.95	0.4650
CM X S/G	(Error Term for CM, G X CM)					
CO X CM	2	1	28	0.9662	0.47	0.6291
G X CO X CM	2	3	28	0.8055	1.03	0.4175
CO X CM X S/G	(Error Term for CO X CM, G X CO X CM)					

\* significant at  $\alpha = 0.05$

where: dy = number of dependent measures

$\frac{df}{H}$  = degrees of freedom for treatment effect

$\frac{df}{E}$  = degrees of freedom for error effect

U = Wilk's likelihood ratio statistic

$$\frac{|E|}{|E + H|}, \text{ where:}$$

$|E|$  = determinant of sum of squares and cross-products for error

$|E + H|$  = determinant of the sum of the sum of squares and cross-products matrix for error, and the sum of squares and cross-products matrix for treatment

TABLE 44

MANOVA Summary Table, Point-Comparison Experiment, for the Dependent Measures of MWLMEAN, Preference

<u>Source</u>	<u>dy</u>	$\frac{df}{H}$	$\frac{df}{E}$	<u>U</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>						
Graph (G)	2	3	28	0.6312	2.33	0.0452 *
Subject (S)/G	(Error Term for G)					
<u>Within-Subject</u>						
Coding (CO)	2	1	28	0.2839	34.05	0.0001 *
G X CO	2	3	28	0.5571	3.06	0.0120 *
CO X S/G	(Error Term for CO, G X CO)					
Complexity (CM)	2	1	28	0.1384	83.98	0.0001 *
G X CM	2	3	28	0.8231	0.92	0.4880
CM X S/G	(Error Term for CM, G X CM)					
CO X CM	2	1	28	0.9310	1.00	0.3814
G X CO X CM	2	3	28	0.9201	0.38	0.8870
CO X CM X S/G	(Error Term for CO X CM, G X CO X CM)					

\* significant at  $\alpha = 0.05$

where: dy = number of dependent measures

$\frac{df}{H}$  = degrees of freedom for treatment effect

$\frac{df}{E}$  = degrees of freedom for error effect

U = Wilk's likelihood ratio statistic

$$\frac{|E|}{|E + H|}, \text{ where:}$$

$|E|$  = determinant of sum of squares and cross-products for error

$|E + H|$  = determinant of the sum of the sum of squares and cross-products matrix for error, and the sum of squares and cross-products matrix for treatment



TABLE 45

Newman-Keuls Analysis, Point-Comparison Experiment, of  
MWLMEAN for the Main Effects of Graph-Type \*

---

GRAPH-TYPE				
GRAPH	POINT	BAR	LINE	3DBAR
MWL- MEAN	4.41	4.95	5.48	5.60

---

\* Treatments underlined by a common line do not differ from each other at  $p \leq 0.05$

TABLE 46

Duncan's Multiple Range Analysis, Point-Comparison  
Experiment, of MWLMEAN for the Main Effects of Graph-Type \*

GRAPH-TYPE				
GRAPH	POINT	BAR	LINE	3DBAR
MWL- MEAN	4.41	4.95	5.48	5.60

\* Treatments underlined by a common line do not differ  
from each other at  $p \leq 0.05$

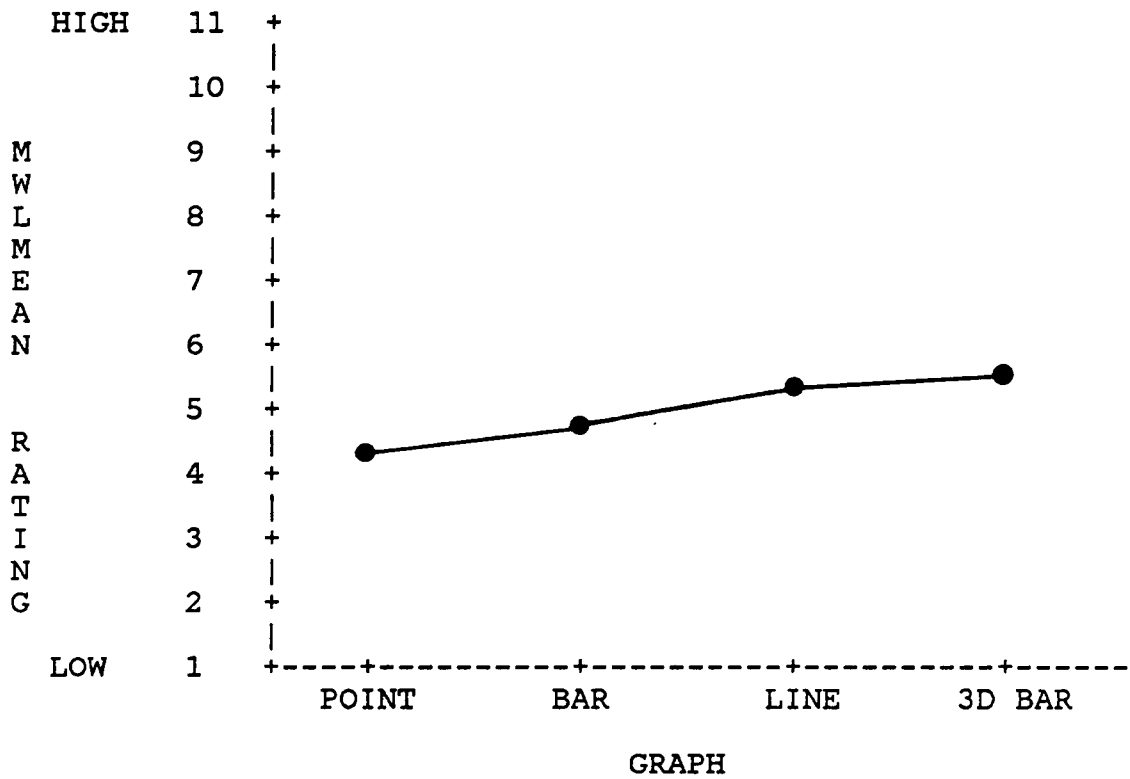


Figure 31. Effect of graph-type, point-comparison experiment, on MWLMEAN rating.

Trend-reading experiment. MANOVA results for objective measures (Table 47) were significant for coding ( $F(2,27)=3.98$ ,  $p=0.0305$ ), complexity ( $F(2,27)=105.11$ ,  $p=0.0001$ ), graph-by-coding ( $F(6,54)=2.51$ ,  $p=0.0324$ ), and a graph-by-coding-by-complexity interaction ( $F(6,54)=2.52$ ,  $p=0.0320$ ) that was not found significant in earlier MANOVA's. Error score ( $F(3,28)=4.69$ ,  $p=0.0089$ ), as indicated in Table 29, was the only measure found significant for this last effect. A Newman-Keuls analysis (Table 48) indicated that high complexity, color-coded, three-dimensional bar graphs had a significantly higher percentage of errors than all other types of graphs, while there was no significant difference between any of the other conditions (Figure 32).

The MANOVA of subjective measures (Table 49) was significant for coding ( $F(2,27)=28.66$ ,  $p=0.0001$ ) and complexity ( $F(2,27)$ ,  $p=0.0001$ ), and was almost significant for the graph-by-complexity interaction ( $F(6,54)=2.19$ ,  $p=0.0584$ ). Therefore, no changes were revealed in this MANOVA in comparison to the previous overall MANOVA on all four dependent measures (Table 11).

TABLE 47

MANOVA Summary Table, Trend-Reading Experiment,  
for the Dependent Measures of Percent Error, Time

<u>Source</u>	<u>dy</u>	<u>df</u> H	<u>df</u> E	<u>U</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>						
Graph (G)	2	3	28	0.7737	1.23	0.3047
Subject (S)/G	(Error Term for G)					
<u>Within-Subject</u>						
Coding (CO)	2	1	28	0.7721	3.98	0.0305 *
G X CO	2	3	28	0.6113	2.51	0.0324 *
CO X S/G	(Error Term for CO, G X CO)					
Complexity (CM)	2	1	28	0.1138	105.11	0.0001 *
G X CM	2	3	28	0.7404	1.46	0.2099
CM X S/G	(Error Term for CM, G X CM)					
CO X CM	2	1	28	0.9530	0.67	0.5223
G X CO X CM	2	3	28	0.6105	2.52	0.0320 *
CO X CM X S/G	(Error Term for CO X CM, G X CO X CM)					

\* significant at  $\alpha = 0.05$

where: dy = number of dependent measures

$\frac{df}{H}$  = degrees of freedom for treatment effect

$\frac{df}{E}$  = degrees of freedom for error effect

U = Wilk's likelihood ratio statistic

$$\frac{|E|}{|E + H|}, \text{ where:}$$

|E| = determinant of sum of squares and cross-products for error

|E + H| = determinant of the sum of the sum of squares and cross-products matrix for error, and the sum of squares and cross-products matrix for treatment

TABLE 48

Newman-Keuls Analysis, Trend-Reading Experiment, of Percent Error Score for the Interactive Effects of Graph-by-Coding-by-Complexity \*

GRAPH CODING COMPLEX	LIN COL LO	LIN B&W LO	PNT COL LO	BAR COL LO	BAR COL HI	3DB COL LO	LIN COL HI
PERCENT ERROR	0.00	0.00	0.00	0.00	0.00	0.00	3.12

(continued)

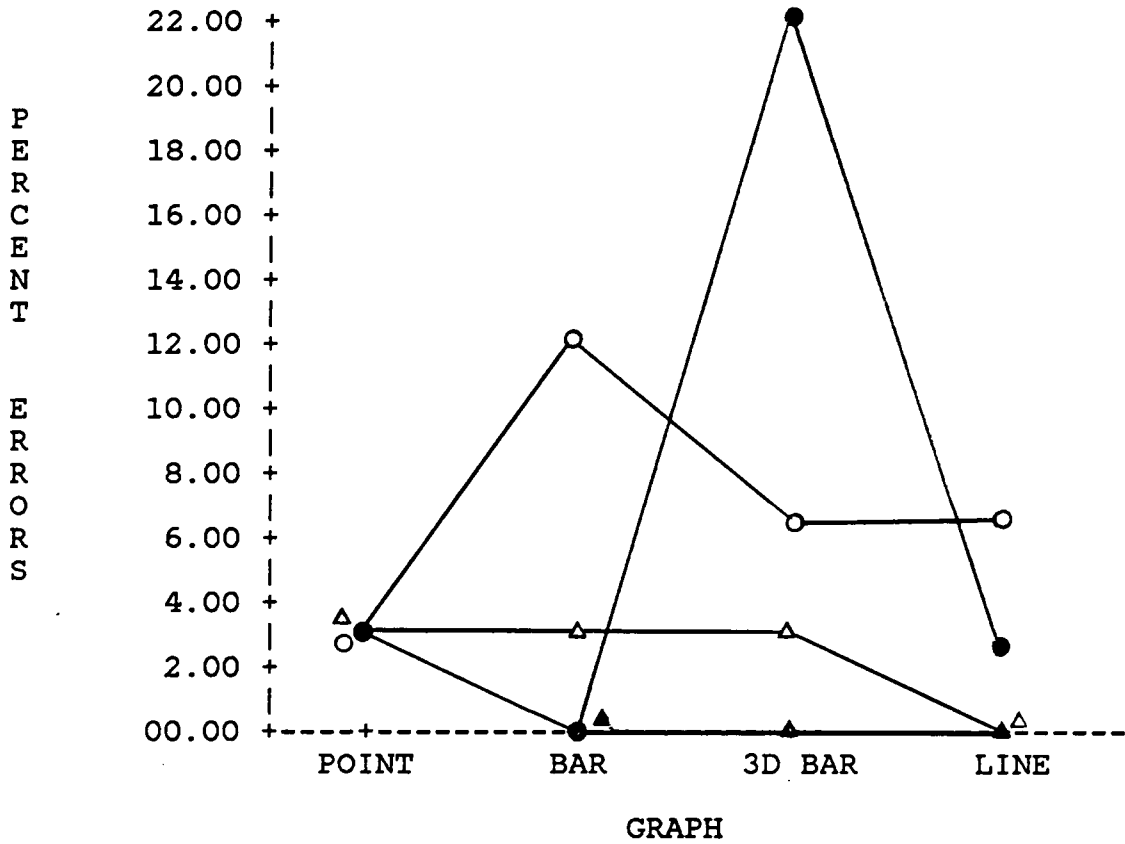
GRAPH CODING COMPLEX	PNT B&W LO	PNT COL HI	PNT B&W HI	BAR B&W LO	3DB B&W LO	LIN B&W HI	3DB B&W HI
PERCENT ERROR	3.12	3.12	3.12	3.12	3.12	6.25	6.25

(continued)

GRAPH CODING COMPLEX	BAR B&W HI	3DB COL HI
PERCENT ERROR	12.50	21.87

\* Treatments underlined by a common line do not differ from each other at  $p \leq 0.05$

where: LIN = Line graphs  
 BAR = Bar graphs  
 PNT = Point-plot graphs  
 3DB = Three-dimensional bar graphs  
 COL = Color-coding  
 B&W = Black-and-white coding  
 LO = Low complexity  
 HI = High complexity



KEY: High complexity color-coding ●  
 High complexity black-and-white coding ○  
 Low complexity color-coding ▲  
 Low complexity black-and-white coding △

Figure 32. Effect of graph-by-coding-by-complexity interaction, trend-reading experiment, on percent error score.

TABLE 49

MANOVA Summary Table, Trend-Reading Experiment, for the Dependent Measures of MWLMEAN, Preference

<u>Source</u>	<u>dy</u>	$\frac{df}{H}$	$\frac{df}{E}$	<u>U</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>						
Graph (G)	2	3	28	0.7722	1.24	0.3000
Subject (S)/G	(Error Term for G)					
<u>Within-Subject</u>						
Coding (CO)	2	1	28	0.3202	28.66	0.0001 *
G X CO	2	3	28	0.6892	1.84	0.1084
CO X S/G	(Error Term for CO, G X CO)					
Complexity (CM)	2	1	28	0.0934	130.95	0.0001 *
G X CM	2	3	28	0.6472	2.19	0.0584
CM X S/G	(Error Term for CM, G X CM)					
CO X CM	2	1	28	0.9598	0.56	0.5751
G X CO X CM	2	3	28	0.7590	1.33	0.2599
CO X CM X S/G	(Error Term for CO X CM, G X CO X CM)					

\* significant at  $\alpha = 0.05$

where: dy = number of dependent measures

$\frac{df}{H}$  = degrees of freedom for treatment effect

$\frac{df}{E}$  = degrees of freedom for error effect

U = Wilk's likelihood ratio statistic

$$\frac{|E|}{|E + H|}, \text{ where:}$$

|E| = determinant of sum of squares and cross-products for error

|E + H| = determinant of the sum of the sum of squares and cross-products matrix for error, and the sum of squares and cross-products matrix for treatment



Trend-comparison experiment. The MANOVA for objective measures for the trend-comparison experiment (Table 50) exhibited significance for coding ( $F(2,27)=3.67$ ,  $p=0.0389$ ), and complexity ( $F(2,27)=110.18$ ,  $p=0.0001$ ).

Results for the MANOVA on the subjective measures (Table 51) also showed significance for complexity ( $F(2,27)=94.87$ ,  $p=0.0001$ ) and coding ( $F(2,27)=37.07$ ,  $p=0.0001$ ).

These results were the same as those obtained in the MANOVA of all four dependent variables for trend-comparison. Again, ANOVA analyses of the effects for these measures were discussed in an earlier section.

TABLE 50

MANOVA Summary Table, Trend-Comparison Experiment, for the Dependent Measures of Percent Error, Time

<u>Source</u>	<u>dv</u>	<u>df</u> H	<u>df</u> E	<u>U</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>						
Graph (G)	2	3	28	0.7753	1.22	0.3101
Subject (S)/G	(Error Term for G)					
<u>Within-Subject</u>						
Coding (CO)	2	1	28	0.7861	3.67	0.0389 *
G X CO	2	3	28	0.8556	0.73	0.6280
CO X S/G	(Error Term for CO, G X CO)					
Complexity (CM)	2	1	28	0.1091	110.18	0.0001 *
G X CM	2	3	28	0.9041	0.47	0.8311
CM X S/G	(Error Term for CM, G X CM)					
CO X CM	2	1	28	0.9774	0.31	0.7346
G X CO X CM	2	3	28	0.9659	0.16	0.9867
CO X CM X S/G	(Error Term for CO X CM, G X CO X CM)					

\* significant at  $\alpha = 0.05$

where: dv = number of dependent measures

$\frac{df}{H}$  = degrees of freedom for treatment effect

$\frac{df}{E}$  = degrees of freedom for error effect

U = Wilk's likelihood ratio statistic

$$\frac{|E|}{|E + H|}, \text{ where:}$$

$|E|$  = determinant of sum of squares and cross-products for error

$|E + H|$  = determinant of the sum of the sum of squares and cross-products matrix for error, and the sum of squares and cross-products matrix for treatment

TABLE 51

MANOVA Summary Table, Trend-Comparison Experiment, for  
the Dependent Measures of MWLMEAN, Preference

<u>Source</u>	<u>dy</u>	<u>df</u> H	<u>df</u> E	<u>U</u>	<u>F</u>	<u>p</u>
<u>Between-Subjects</u>						
Graph (G)	2	2	28	0.7693	1.26	0.2909
Subject (S)/G	(Error Term for G)					
<u>Within-Subject</u>						
Coding (CO)	2	2	28	0.2669	37.07	0.0001 *
G X CO	2	6	28	0.7596	1.33	0.2615
CO X S/G	(Error Term for CO, G X CO)					
Complexity (CM)	2	2	28	0.1245	94.87	0.0001 *
G X CM	2	6	28	0.6902	1.83	0.1099
CM X S/G	(Error Term for CM, G X CM)					
CO X CM	2	2	28	0.8182	3.00	0.0667
G X CO X CM	2	6	28	0.7646	1.29	0.2765
CO X CM X S/G	(Error Term for CO X CM, G X CO X CM)					

\* significant at  $\alpha = 0.05$

where: dy = number of dependent measures

df<sub>H</sub> = degrees of freedom for treatment effect

df<sub>E</sub> = degrees of freedom for error effect

U = Wilk's likelihood ratio statistic

$$\frac{|E|}{|E + H|}, \text{ where:}$$

|E| = determinant of sum of squares and cross-products for error

|E + H| = determinant of the sum of the sum of squares and cross-products matrix for error, and the sum of squares and cross-products matrix for treatment

## DISCUSSION AND CONCLUSIONS

The lack of much current and pertinent human factors data in the field of computer-generated graphics dictated an exploratory, rather than a theoretically-based approach in this study. Four core tasks were chosen for investigation in the four individual experiments. Each task required a different graph reading skill that was felt to be a necessary component of interpreting graphically-presented data. In each task, four graph-types, two levels of task complexity, and two levels of coding were investigated in relation to their effects on errors, time to complete the task, subjective mental workload, and user preference ratings.

### Main Effects of Complexity

Complexity level was manipulated to determine if any interaction effects with graph-type or coding would result. In this respect, the main effect of complexity was important to this study only so far as determining whether two levels of difficulty were indeed achieved and conveyed to the subjects. As indicated by MANOVA and ANOVA analyses, the two levels of complexity were significantly different for all measures in all tasks except RMS error in the point-reading experiment. This finding strongly

evidenced the presence of actual complexity changes in the task. However, no graph-by-complexity interactions were found.

### Main Effects of Graph-Type

Within each task, four different types of graphs were chosen. Due to the differences in the way these graphs visually represented data, it was felt that they might be expected to account for a large amount of variation in the dependent measures. In general, it was found that this was not the case. Only for the point-reading task were significant differences between graph-types found, and then only for the dependent measure of time to complete the task. Point-reading tasks using three-dimensional bar graphs required more time to complete than line, bar, and point-plot graphs ( $p < 0.05$ ), however there were no statistically-significant differences between the latter three graph-types. A MANOVA of subjective measures did find statistical significance for the main effects of graph-type ( $p < 0.05$ ) in the point-comparison experiment. Further ANOVA analysis indicated significance of this effect only for the dependent measure of MWLMEAN (Table 23). Newman-Keuls analysis indicated no statistically-significant differences between individual graph-types for MWLMEAN. However, analysis with the Duncan Multiple Range

test indicated that point-plot graphs had significantly lower MWLMEAN rating scores than line or three-dimensional bar graphs. Table 52 shows the cell means for each graph-type across all tasks. Graphs in each measure were ranked from the best at top (lowest time, errors, MWLMEAN rating, highest preference rating) to worst at bottom (highest time, errors, MWLMEAN, lowest preference rating). It is emphasized that the rank orders are provided for informative purposes only and interpretation of statistically-significant order effects from them is not possible.

TABLE 52

Cell Means for Graph-Type for all Measures and Tasks

POINT READING		POINT COMPARISON		TREND READING		TREND COMPARISON	
----- ERRORS -----							
BAR	0.480	PNT	9.37	LIN	2.34	PNT	15.62
LIN	0.577	BAR	15.62	PNT	2.34	BAR	20.31
3DB	0.959	3DB	20.31	BAR	3.90	LIN	25.78
PNT	1.466	LIN	20.31	3DB	7.81	3DB	31.25
----- TIME -----							
BAR	18.68	PNT	48.36	PNT	37.20	LIN	70.66
LIN	19.54	BAR	52.39	LIN	39.41	BAR	79.31
PNT	20.28	LIN	54.10	BAR	40.22	PNT	80.41
3DB	35.73	3DB	66.24	3DB	48.90	3DB	81.69
----- MWLMEAN -----							
PNT	3.38	PNT	4.41	PNT	3.40	PNT	5.66
LIN	3.73	BAR	4.95	LIN	3.82	LIN	6.14
BAR	3.77	LIN	5.48	BAR	3.83	BAR	6.18
3DB	4.85	3DB	5.60	3DB	4.24	3DB	6.23
----- PREFERENCE -----							
BAR	6.46	BAR	6.12	BAR	7.07	BAR	5.86
3DB	5.85	3DB	5.41	3DB	7.02	3DB	5.07
LIN	5.06	PNT	5.16	LIN	6.32	LIN	4.63
PNT	5.05	LIN	4.87	PNT	5.71	PNT	4.17

where: LIN = Line graphs  
PNT = Point-plot graphs  
BAR = Bar graphs  
3DB = Three-dimensional bar graphs

It is informative to look at the performance of the different graph types across all tasks, even though statistical comparisons between tasks were not possible (and thus may be highly suspect). Note that tasks using three-dimensional bar graphs consistently took the longest to complete, were rated as requiring the most mean mental effort, had the highest number of errors for two of the tasks, and the second highest for the other two tasks, but were also the second most preferred type of graph. The contradiction in these results may be due to the novelty-appeal of the three-dimensional bar graphs. Most of the subjects who received this type of graph indicated that they had rarely, or never encountered it prior to this study.

#### Main Effects of Coding

The effect of coding was consistent, if not always statistically-significant across all measures for all tasks. Statistically-significant results were found for preference ratings in all experiments, and for time to complete the task and MWLMEAN in the point-comparison, trend-reading, and trend-comparison experiments. The dependent measure of error was not significant in any of the tasks. Tasks using graphs with color-coding required



less time, were rated as requiring less mean mental effort, and were preferred over black-and-white coding. It is clear from these results that color-coding tended to be more effective than black-and-white coding for the types of tasks investigated in this study. A summary of the cell means of for the coding variable are presented in Table 53. Again, rank orders are not statistically-significant, and caution is advised when using this table.

TABLE 53

Cell Means for Coding for all Measures and Tasks

---

POINT READING		POINT COMPARISON		TREND READING		TREND COMPARISON	
----- ERRORS -----							
COL	0.831	COL	14.06	COL	3.51	COL	22.65
B&W	0.909	B&W	18.75	B&W	4.68	B&W	23.82
----- TIME -----							
COL	23.46	COL	52.02	COL	39.29	COL	74.79
B&W	23.65	B&W	58.52	B&W	43.61	B&W	81.24
----- MWLMEAN -----							
COL	3.86	COL	4.87	COL	3.60	COL	5.79
B&W	4.01	B&W	5.35	B&W	4.05	B&W	6.31
----- PREFERENCE -----							
COL	6.73	COL	6.62	COL	7.75	COL	6.05
B&W	4.49	B&W	4.16	B&W	5.31	B&W	3.82

---

where: COL = Color-coding  
 B&W = Black-and-white coding

### Graph-by-Coding Interaction

The presence of a graph-by-coding interaction in three of the tasks indicated that conclusions based solely on the two main effects of graph-type and coding would not fully describe the data. It should be noted however, that statistically-significant interaction effects were found for a total of only three measures within all experiments.

In the point-reading experiment, black-and-white point-plot graphs had the highest Root Mean Square (RMS) error scores, while black-and-white bar, and line graphs had the lowest, and second lowest error scores, respectively. Statistically-significant differences were found only for these two extreme sets of scores (again, see Table 20, and Figure 7).

The point-comparison experiment showed a graph-by-complexity interaction for the mental workload mean (MWLMEAN) measures. In general, color and black-and-white coded point-plot graphs were rated as requiring the least mean mental effort, while black-and-white line and three-dimensional bar graphs had the highest ratings. However, statistically-significant differences between treatment condition means were of a complex nature, and were not found between all combinations of graph-type and coding (refer to Table 27, and Figure 15).

Lastly, the trend-reading experiment exhibited a graph-by-coding interaction for the dependent measure of percent error score. Color-coded, three-dimensional bar graphs, which had the highest percent error score, were found to be significantly different from both color-coded bar graphs, which had the lowest errors, and line and point-plot graphs, which had the second lowest percent error scores (Table 34, and Figure 23). The cell means for the graph-by-coding interaction are shown in Table 54, ranked from best to worst for each dependent measure.

TABLE 54

Cell Means for Combinations of Graph-Type and Coding for all Measures and Tasks

POINT READING		POINT COMPARISON		TREND READING		TREND COMPARISON	
----- ERRORS -----							
BAR		PNT		BAR		PNT	
B&W	0.158	COL	9.37	COL	0.00	COL	14.06
LIN		PNT		PNT		BAR	
B&W	0.310	B&W	9.37	COL	1.56	COL	15.62
PNT		LIN		LIN		PNT	
COL	0.742	COL	14.06	COL	1.56	B&W	17.18
BAR		BAR		LIN		LIN	
COL	0.801	COL	15.62	B&W	3.12	B&W	23.43
LIN		BAR		PNT		BAR	
COL	0.844	B&W	15.62	B&W	3.12	B&W	25.00
3DB		3DB		3DB		LIN	
COL	0.937	COL	17.18	B&W	4.68	COL	28.12
3DB		3DB		BAR		3DB	
B&W	0.980	B&W	23.43	B&W	7.81	B&W	29.68
PNT		LIN		3DB		3DB	
B&W	2.191	B&W	26.56	COL	10.93	COL	32.81
----- TIME -----							
BAR		BAR		PNT		LIN	
COL	18.22	COL	47.99	B&W	36.80	COL	68.72
LIN		PNT		BAR		LIN	
COL	18.47	B&W	48.06	COL	37.11	B&W	72.59
BAR		PNT		LIN		BAR	
B&W	19.14	COL	48.66	COL	37.25	COL	73.78
PNT		LIN		PNT		PNT	
COL	19.72	COL	50.34	COL	37.59	COL	78.05
LIN		BAR		LIN		3DB	
B&W	20.60	B&W	56.80	B&W	41.56	COL	79.63
PNT		LIN		BAR		PNT	
B&W	20.83	B&W	57.86	B&W	43.33	B&W	82.76
3DB		3DB		3DB		3DB	
B&W	34.04	COL	61.11	COL	45.19	B&W	83.76
3DB		3DB		3DB		BAR	
COL	37.41	B&W	71.37	B&W	52.76	B&W	85.84

where: COL = Color-coding  
 B&W = Black-and-white coding  
 LIN = Line graphs  
 PNT = Point-plot graphs  
 BAR = Bar graphs  
 3DB = Three-dimensional bar graphs

TABLE 54 (continued)

Cell Means for Combinations of Graph-Type and Coding for all Measures and Tasks

POINT READING		POINT COMPARISON		TREND READING		TREND COMPARISON	
----- MWLMEAN -----							
PNT		PNT		PNT		PNT	
COL	3.33	COL	4.38	COL	3.33	COL	5.52
PNT		PNT		PNT		PNT	
B&W	3.43	B&W	4.45	B&W	3.48	B&W	5.80
BAR		BAR		BAR		BAR	
B&W	3.65	COL	4.82	COL	3.62	COL	5.84
LIN		3DB		3DB		3DB	
COL	3.67	COL	5.08	COL	3.72	COL	5.06
BAR		BAR		LIN		LIN	
COL	3.81	B&W	5.09	COL	3.74	COL	5.93
LIN		LIN		LIN		LIN	
B&W	3.86	COL	5.21	B&W	3.90	B&W	6.35
3DB		LIN		BAR		BAR	
COL	4.60	B&W	5.75	B&W	4.05	B&W	6.25
3DB		3DB		3DB		3DB	
B&W	5.10	B&W	6.12	B&W	4.77	B&W	6.59
----- PREFERENCE -----							
BAR		BAR		BAR		BAR	
COL	8.12	COL	7.79	COL	8.71	COL	7.39
3DB		3DB		3DB		3DB	
COL	6.56	COL	6.59	COL	8.14	COL	6.12
LIN		LIN		LIN		LIN	
COL	6.45	COL	6.26	COL	7.68	COL	5.92
PNT		PNT		PNT		PNT	
COL	5.78	COL	5.82	COL	6.46	COL	4.78
3DB		PNT		3DB		BAR	
B&W	5.15	B&W	4.50	B&W	5.90	B&W	4.34
BAR		BAR		BAR		3DB	
B&W	4.81	B&W	4.45	B&W	5.43	B&W	4.01
PNT		3DB		PNT		PNT	
B&W	4.32	B&W	4.23	B&W	4.95	B&W	3.57
LIN		LIN		LIN		LIN	
B&W	3.67	B&W	3.48	B&W	4.95	B&W	3.34

where: COL = Color-coding  
 B&W = Black-and-white coding  
 LIN = Line graphs  
 PNT = Point-plot graphs  
 BAR = Bar graphs  
 3DB = Three-dimensional bar graphs

### Graph-by-Complexity-by-Coding Interaction

A three-way interaction effect was found for the variables of graph-type, complexity, and coding in the trend-reading experiment. High complexity, color-coded, three-dimensional bar graphs had significantly higher percent error scores than all other conditions (Table 48, and Figure 32).

### Conclusions and Recommendations

Several important conclusions can be drawn regarding the independent variables investigated in this study.

First, the results of this study indicate that more time is needed to complete point-reading tasks with three-dimensional bar graphs than either line, bar, or point-plot graphs. These results were somewhat predictable considering the difficulty of locating specific points on the three-dimensional bar graph because the data values are represented both by positioning across the width of the screen, and apparent depth into the screen. In addition, depth perception cues normally present in nature, such as accommodation differences between objects at different distances, and perspective (i.e. graphs were isometric and had no vanishing point), were missing in the three-dimensional bar graphs. Differences in time to complete

the task between line, vertical-bar, and horizontal-bar graphs found by Schutz (1969a) in a task similar to the trend-reading experiment of this study, were not exhibited.

Due to the large apparent differences between graph-types it was expected that statistically-significant results would also be found among some of the other graphs. While the line, point-plot, and bar graphs were all two-dimensional representations, differences among the graphs on the four experimental tasks were still expected because they differed on other important qualities such as the method of point connection, and type of overlap of data points when more than one data set was represented. For example, one might have expected that line graphs, the only graph to use a "peak to peak" method of point connection (points were connected by a line), would be better than other types of graphs in the trend-reading, and trend-comparison experiments. This is because line graphs depict differences in the values between connected points by differences in the angle of the connecting line as well as perceived height above the X-axis, thus adding a certain amount of redundancy. It may well be, however, that the visual clutter produced by connecting points with lines was distracting, or that area cues provided by bars, in the bar graphs, were more effective than was earlier expected. In addition, because line graphs were one of the graphical



techniques that did not use a method that connected points by a bar to the bottom of the graph, it may have been more difficult to exactly locate an X-axis referenced point on a line. Using the same reasoning, bar graphs should have been easier to use in the point-reading and point-comparison tasks where location of specific X-axis referenced points was necessary. Point-plot graphs, with no point connection method at all would have presumably been the the worst of the three types of graphs for all tasks except point-reading. Again, however, the amount of visual clutter on point-plot graphs was less than that of the other three graphs and this may have led to reduced confusion and visual search time.

The effects of inherent graph complexity were strongly evidenced. Recall that Schutz (1961b) in a task much like the point-reading task of this study also found that when fewer data sets, and data points were used, less time was needed to complete the task. In addition, he found that when less graph lines were crossed (low confusion), less time was needed to complete the task than when many lines were crossed (high confusion). These results are similar in nature to those found for the point-reading experiment in this study, mainly that high complexity graphs took longer to complete, and were rated higher on MWLMEAN, and lower on preference.

Results for the other three experiments also indicate that limiting the amount of information on a graph, if possible, may be preferred, as well as reduce task time, errors, and subjective mental effort. Because the complexity of the question and the associated graph were both varied for these latter three experiments (point-comparison, trend-reading, and trend-comparison), it is not possible to determine exactly where the locus of the complexity effect resides. That is, whether increased complexity of the graph, or of the question was responsible for variations in the dependent measures. Even though the locus of this effect was not one of the prime objectives of the study, these results show that the combination of high complexity graphs and questions is to be avoided if at all possible.

The prominent effects found between the color, and black-and-white coding conditions are similar to the results from past research. In particular, color-coding reduced task time, a result also found by Aretz and Calhoun (1982), and Schutz (1961b). However no statistically-significant differences in error scores were found between color, and black-and-white coding. These results appear to indicate that while color-coding reduces visual search time, the differences between color, and black-and-white

coding are not so great as to cause significant differences in misrecognition and graph interpretation errors. Where time considerations are of importance, such as in a lecture or presentation, color-coding may reduce the time that is necessary to convey graphically-presented information. The similar findings of decreased MWLMEAN for color-coding indicate that persons may perceive graphs encoded this way as being less complex and less difficult to understand. This may be important in situations where impressions of clarity, and freedom from complexity are essential, such as in presentations to non-technical personnel, or as in advertisement. Graphics that give the impression of being extremely complex and difficult at first glance might also influence whether a person will invest the time and effort involved in investigating the relationships shown. It may even be the case that simple relationships could be made to seem unnaturally complex by the use of an inappropriate graphing technique.

Color-coding was preferred over black-and-white coding in all of the experiments. These results were also found in the earlier studies of Tullis (1981) and Schutz (1961b). This is an important result when attention-getting and attention-maintaining ability is considered. In instances where large amounts of information are being presented, color-coding may help to initially draw a person's

attention to a graph, or to a specific detail of a graph. It may also indicate to the viewer, as evidenced by its high preference rating in this study, that more time and effort has been expended to produce a quality presentation. If it is strongly desired to influence the viewer, color-coding may be more appropriate than black-and-white coding due to possible crossover effects from the method of presentation to the topic of the presentation.

Before one totally embraces color-coding as the method of choice over black-and-white, two prominent factors should be considered. First, it is more difficult and expensive to reproduce color-coded information in hardcopy. It may not be possible to duplicate on hardcopy the hues, saturations, and brightnesses of specific colors displayed on the VDU. Furthermore the use of color-coding may cause problems with color-blind individuals, as they may be unable to discriminate between certain hues chosen for a coding scheme. This problem, of course, does not exist with black-and-white coding systems, although other visual problems, such as poor minimum separable acuity, may make it difficult for the viewer to discern black-and-white cross-hatching.

The three graph-by-coding interactions give somewhat inconclusive, but useful information regarding certain

graph parameters and graph-reading tasks. For the point-reading experiment, black-and-white bar and line graphs produced the lowest, and second lowest RMS error scores respectively, while black-and-white point-plot graphs are to be avoided because they resulted in the highest RMS error scores. For the point-comparison experiment, a graph-by-coding interaction indicated that color and black-and-white coded point-plot graphs were rated as requiring the least mean mental effort, while black-and-white line, and three-dimensional bar graphs had the highest ratings and should also be avoided. However, as was discussed earlier, the statistically-significant differences between the treatment means of this interaction were of a complex nature, and were not found between all treatment conditions. Finally, in the trend-reading experiment, color-coded bar, point-plot, and line graphs produced the fewest errors, while color-coded three-dimensional bar graphs had the greatest numbers of errors and thus probably should not be used for this task.

Aside from indicating a few combinations that should be avoided under certain circumstances, the above findings are not supported by enough significant effects to come to any overall conclusions about the graph-by-coding interaction. That this interaction was statistically significant in at least a few cases is not surprising

considering the way these two variables affect each other. For example, the use of color in a bar graph requires the filling of a relatively large area (the bar interior) with color, while in a point-plot graph that area was only a small symbol. That is, the addition of color-coding in this specific example does not produce equal amounts of colored area on the two types of graphs. It cannot be fully determined from this data whether the addition of color-coding to small and hard to find symbols or points is as effective as its addition is to large areas.

Lastly, the three-way, graph-by-coding-by-complexity interaction indicated that high complexity, color-coded, three-dimensional bar graphs had significantly higher percent error scores than all other conditions for the trend-reading experiment, and thus should not be used for this type of task if errors are to be minimized.

The above findings indicate that further research into the effects of different types and variations of coding techniques is needed. In particular, the major differences between graphs for particular tasks were not found, while coding, which earlier had been felt would only produce subtle differences, had major effects upon some of the dependent measures. These variables need to be further addressed with different types of graphs for different

types of interpretation tasks.

## REFERENCES

- Aretz, A. J. and Calhoun, G. L. Computer generated pictorial stores management displays for fighter aircraft. Proceedings of the Human Factors Society 26th Annual Meeting. Santa Monica, CA: Human Factors Society, 1982, 455-459.
- Barkin, S. An investigation into some factors affecting information system utilization. Unpublished Ph.D. thesis, University of Minnesota, 1971 (in Dickson, Senn and Chervany, 1977, 919-920).
- Benbasat, I. and Schroeder, R. G. An experimental investigation of some MIS design variables. The Management Information Systems Quarterly, 1977, (in Lucas, 1981, 758)
- Burchi, R. S. Interactive graphics today. IBM Systems Journal, 1980, 19, 292-313.
- Casali, J. G. A sensitivity/intrusion comparison of mental workload estimation techniques using a simulated flight task emphasizing perceptual piloting behaviors, Ph.D. Dissertation, Virginia Polytechnic Institute and State University, 1982.
- Casali, J. G. and Wierwille, W. W. Effects on fourteen workload metrics of variations in pilot perceptual workload in a simulated flight emphasizing perceptual activity. Ergonomics, 1984, 27, 1033-1050.
- Casali, J. G. and Wierwille, W. W. A comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. Human Factors, 1983, 6, 623-641.
- Christman, R. J. Sensory Experience. New York: Intext Educational Publishers, 1971.
- Dickson, G. W., Chervany, N. L. and Kozar, L. A. An experimental gaming framework for investigating the influence of management information systems on decision effectiveness. Minneapolis, Minnesota: University of Minnesota, Management Information Working Paper Series, MISRC-WP-72-12, April, 1972 (in Dickson, Senn and Chervany, 1977, 913-914).



Dickson, G. W., Senn, J. A. and Chervany, N. L. Research in management information systems: the Minnesota experiments. Management Science, 1977, 23, 913-923.

Finkelman, J. M., Wolf, E. W. and Friend, M. A. Modified discriminant analysis as a multivariate post-comparison extension of MANOVA for interpretation of simultaneous multimodality measures. Human Factors, 1977, 19, 253-261.

IBM Corporation, Basic (second edition), IBM, May 1982.

IBM Corporation, DOS (second edition), IBM, May 1982.

Koch, C. G., Edman, T. R. and Guenther, K. R. Evaluating options for CRT display of process trend data. Proceedings of the Human Factors Society 26th Annual Meeting. Santa Monica, CA: Human Factors Society, 1982, 49-53.

Lockwood, A. Diagrams: A visual survey of graphs, maps, charts, and diagrams, Studio Vista, 1969.

Lucas, H. C. Jr. An experimental investigation of the use of computer-based graphics in decision making. Management Science, 1982, 27, 757-768.

Lucas, H. C. and Nielson, N. R. The impact of the mode of information presentation on learning and performance. Management Science, 1980, 26, 982-993.

Machover, C. Graphic displays. IEEE Spectrum, 1977, 14, 24-32.

Milroy, R. and Poulton, E. C., Labeling graphs for improved reading speed. Ergonomics, 1978, 21, 55-61.

National Research Council, Committee on Vision, Video Displays, Work, and Vision, National Academy Press, Washington, D.C., 1983, 3.

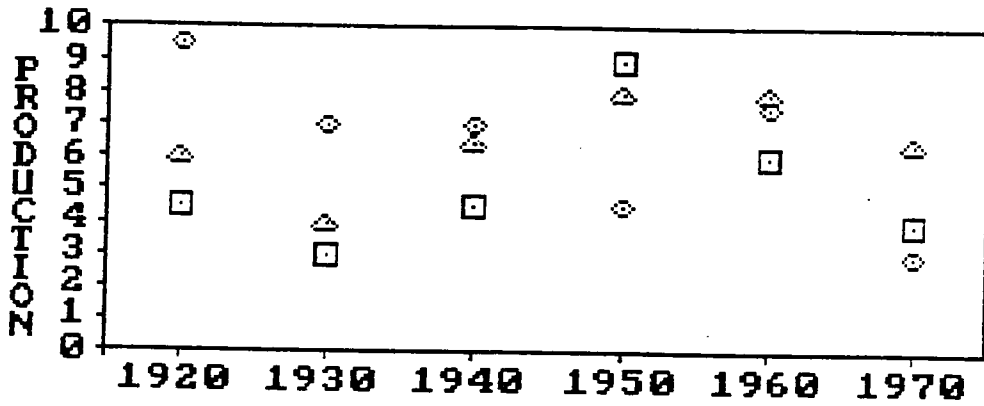
Rahimi, M., and Wierwille, W. W. Evaluation of the sensitivity and intrusion of workload estimation techniques in piloting tasks emphasizing mediational activity. In Proceedings of the 1982 IEEE International Conference on Cybernetics and Society. New York: IEEE 1982, 593-597.

SAS Institute Inc., Statistical Analysis System, 1982a.

- SAS Institute Inc., SAS Users Guide: Statistics. 1982b, 11.
- Schutz, H. G. An evaluation of formats for graphics trend displays - experiment II. Human Factors, 1961 (July), 99-107. (a)
- Schutz, H. G. An evaluation of methods for presentation of graphics multiple trends - experiment III. Human Factors, 1961 (July), 108-119. (b)
- Scott, B. and Wickens, C. D. Spatial and verbal displays in a C3 probabilistic information integration task. Proceedings of the Human Factors Society 27th Annual Meeting. Santa Monica, CA: Human Factors Society, 1983, 355-358.
- Smith, H. R. Experimental comparison of database inquiry techniques, Unpublished Ph.D. Thesis, University of Minnesota, 1975 (in Dickson, Senn and Chervany, 1977, 919).
- Tullis, T. S. An evaluation of alphanumeric, graphic, and color information displays. Human Factors, 1981, 23, 541-550.
- Verhagen, L. J. M. Experiments with bar graph process supervision displays on VDU's. Applied Ergonomics, 1981, 3, 39-45.
- Zelazney, G. and Roche, S. Choosing and using charts. Mckinney and Company, 1972, 1-40.

APPENDIX I

Examples of Experimental Tasks



□ OATS

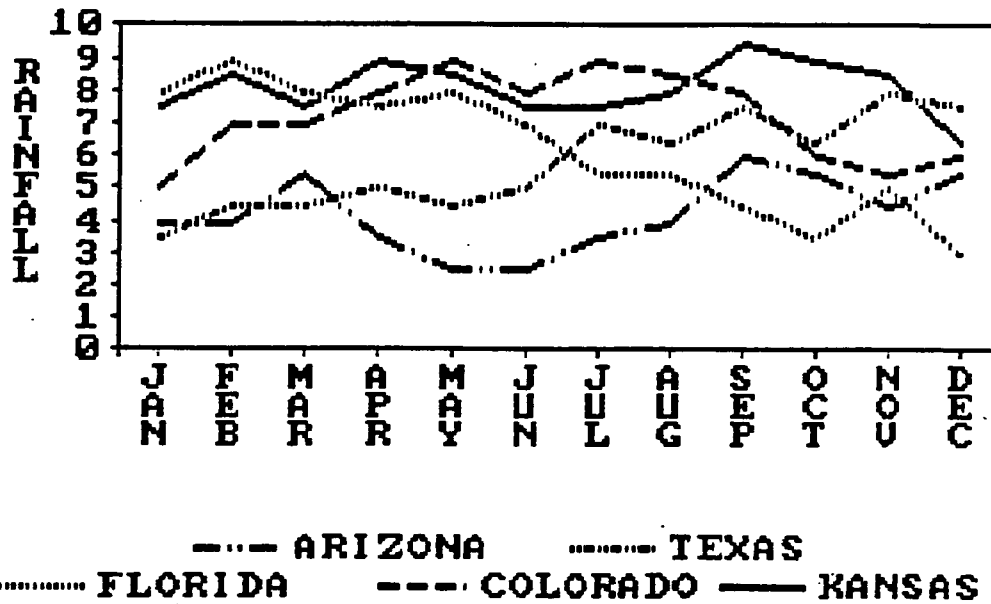
△ HOPS

○ BARLEY

How much Hops was produced in 1940 ?

Enter the correct number: \_\_\_\_\_

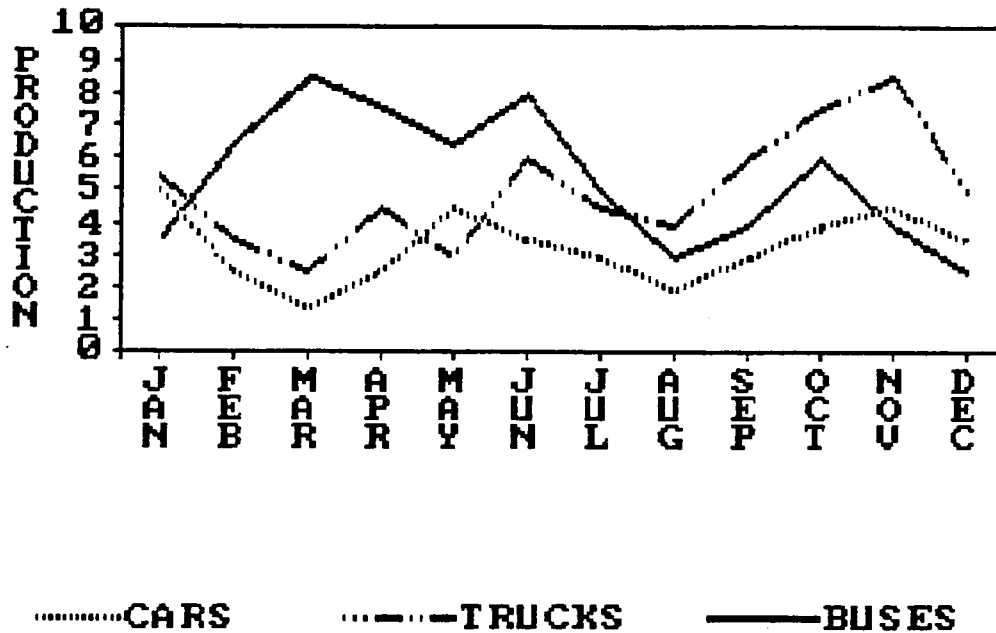
Example of a point-reading task, question and graph pair.  
Variation of the complexity level of this type of task was  
not possible (Scale: 1 cm = 1.69 cm).



What was the amount of rainfall for Kansas in March ?

Enter the correct number: \_\_\_\_\_

Example of a point-reading task, question and graph pair. Variation of the complexity level of this type of task was not possible (Scale: 1 cm = 1.69 cm).

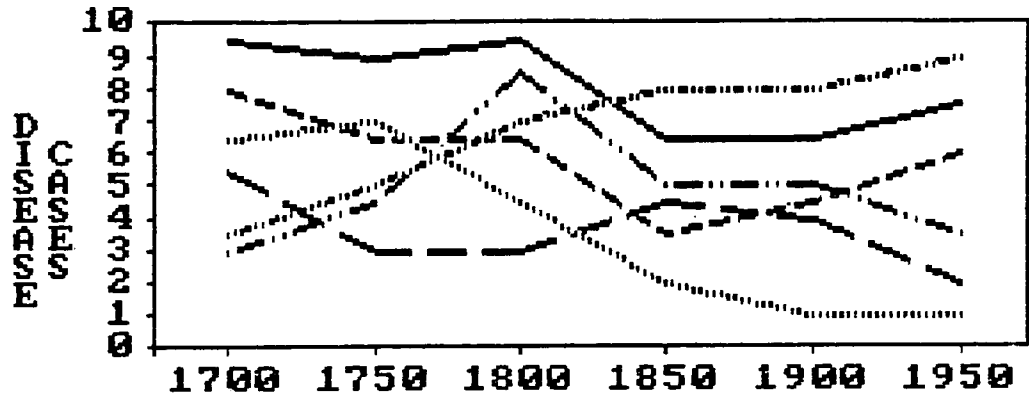


Which had the highest production ?

- F1 Cars in January
- F2 Trucks in October
- F3 Cars in May
- F4 Buses in August

Enter the correct function key:\_\_\_

Example of a point-comparison task, low complexity question and graph pair (Scale: 1 cm = 1.69 cm).



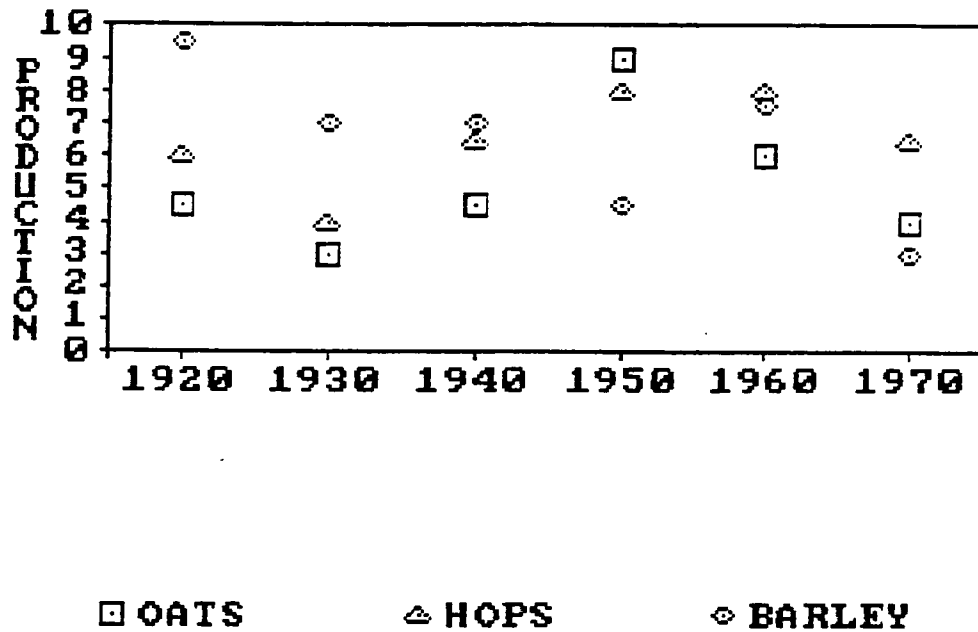
- - - - - FOLIO      - - - - - SMALLPOX      ..... MUMPS  
 ..... PLAGUE      - - - - - CHICKENPOX      ——— MEASLES

Which had the third highest cases of disease ?

- F1 Plague in 1800
- F2 Mumps in 1700
- F3 Smallpox in 1900
- F4 Chickenpox in 1750

Enter the correct function key:\_\_\_

Example of a point-comparison task, high complexity question and graph pair (Scale: 1 cm = 1.69 cm).



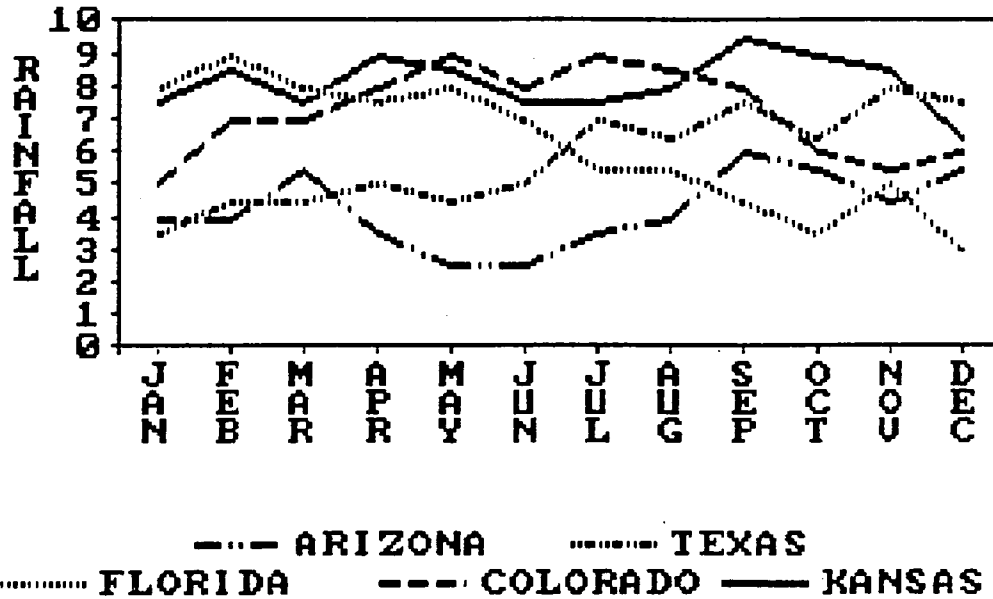
Which is decreasing in production ?

- F1 Hops from 1950 to 1960
- F2 Barley from 1950 to 1960
- F3 Barley from 1930 to 1940
- F4 Oats from 1960 to 1970

Enter the correct function key:\_\_\_

Example of a trend-reading task, low complexity question and graph pair (Scale: 1 cm = 1.69 cm).



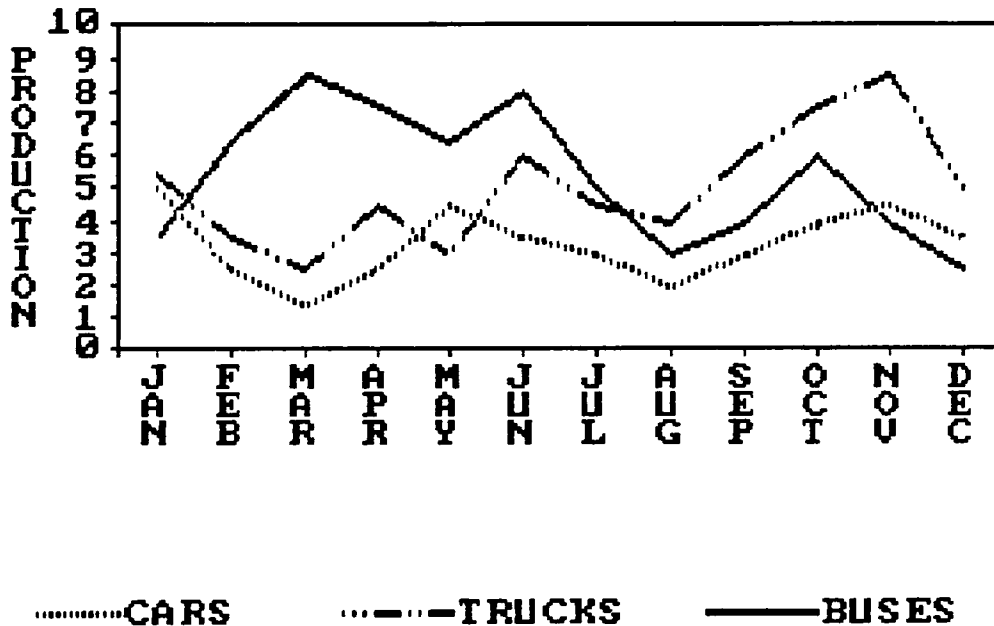


Which is increasing in rainfall at first, then decreasing, and then increasing ?

- F1 Florida from February to June
- F2 Texas from January to May
- F3 Colorado from March to July
- F4 Kansas from June to October

Enter the correct function key:\_\_\_

Example of a trend-reading task, high complexity question and graph pair (Scale: 1 cm = 1.69 cm).

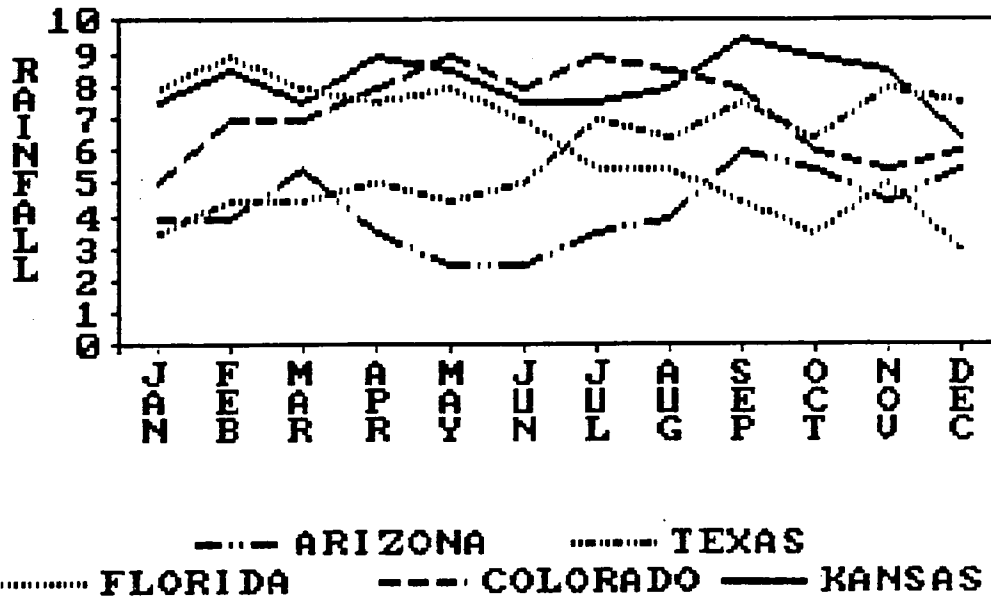


Which decreased the most in production ?

- F1 Trucks from January to March
- F2 Cars from May to July
- F3 Buses from October to December
- F4 Buses from March to May

Enter the correct function key:\_\_\_

Example of a trend-comparison task, low complexity question and graph pair (Scale: 1 cm = 1.69 cm).



Which had the same increases and decreases in rainfall as Arizona from October to December ?

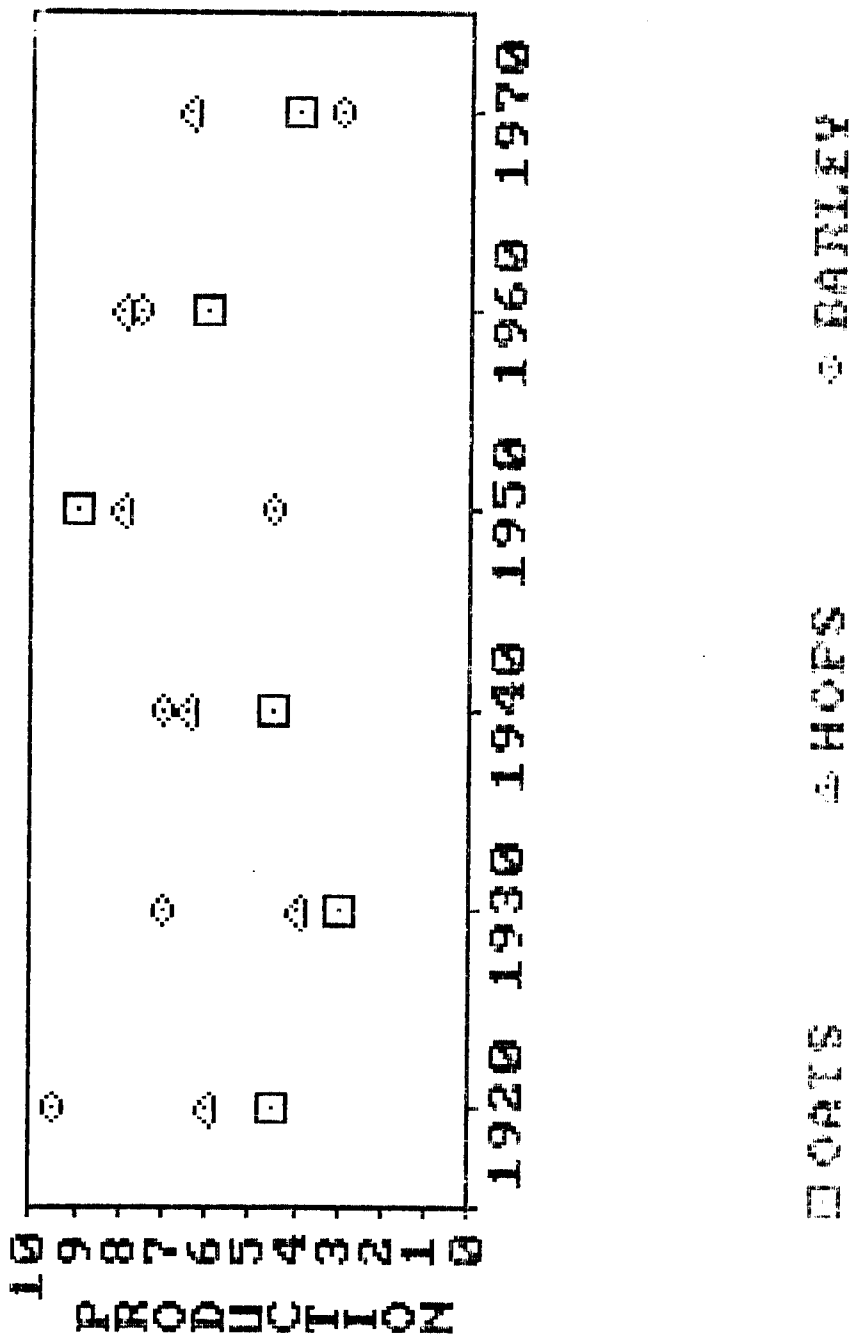
- F1 Texas from August to October
- F2 Colorado from May to July
- F3 Florida from February to April
- F4 Kansas from April to June

Enter the correct function key:\_\_\_

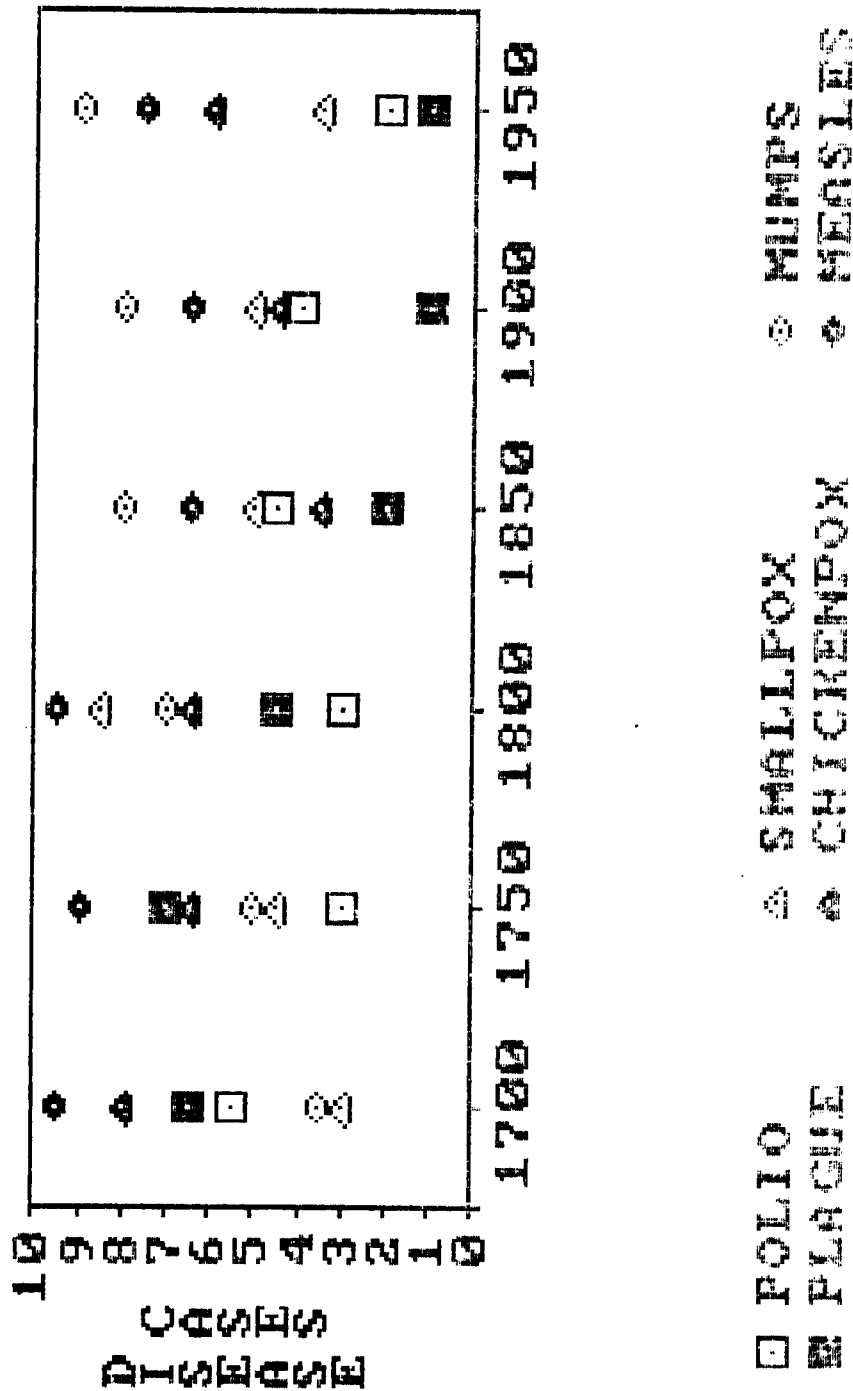
Example of a trend-comparison task, high complexity question and graph pair (Scale: 1 cm = 1.69 cm).

APPENDIX II

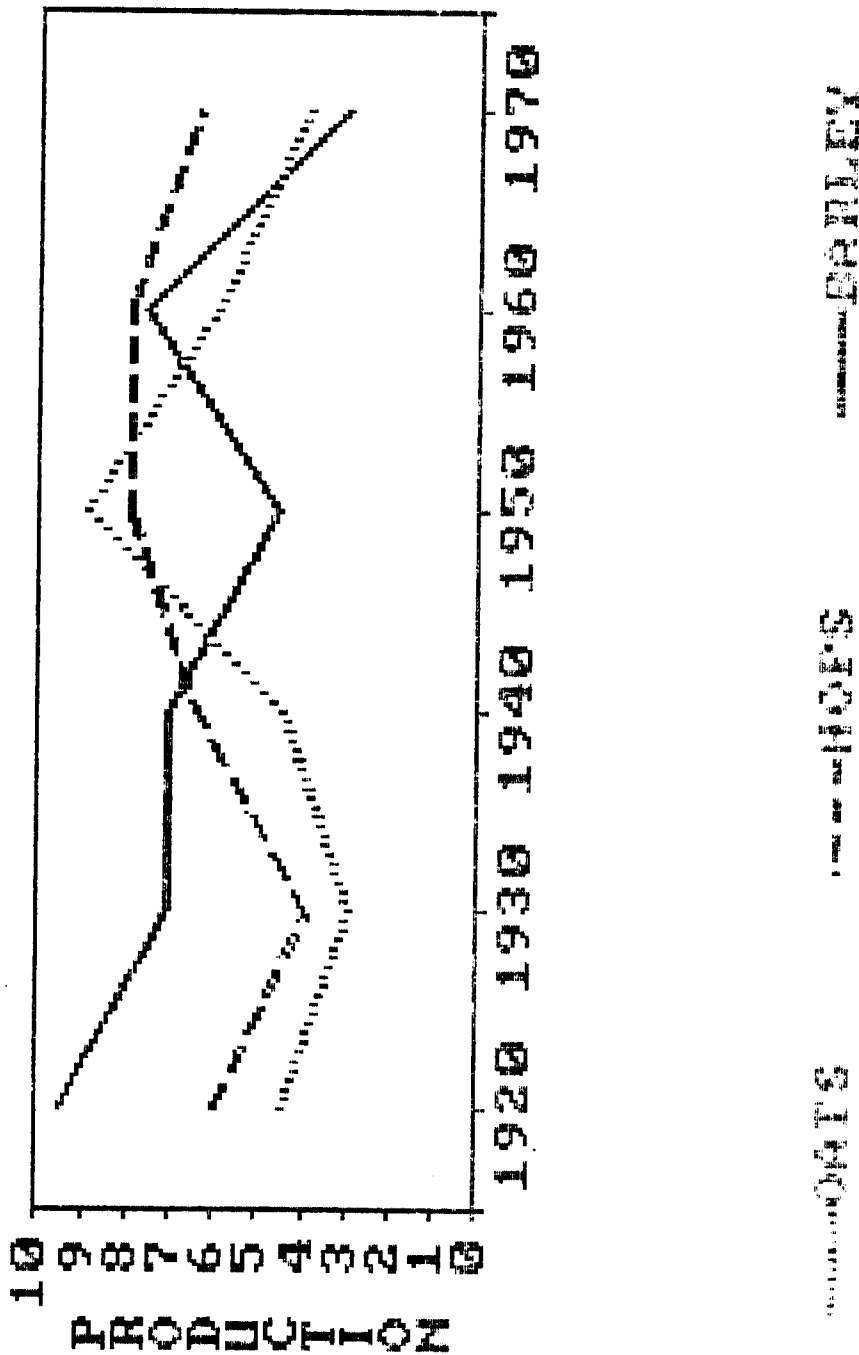
Examples of Black-and-White Coded Graphs



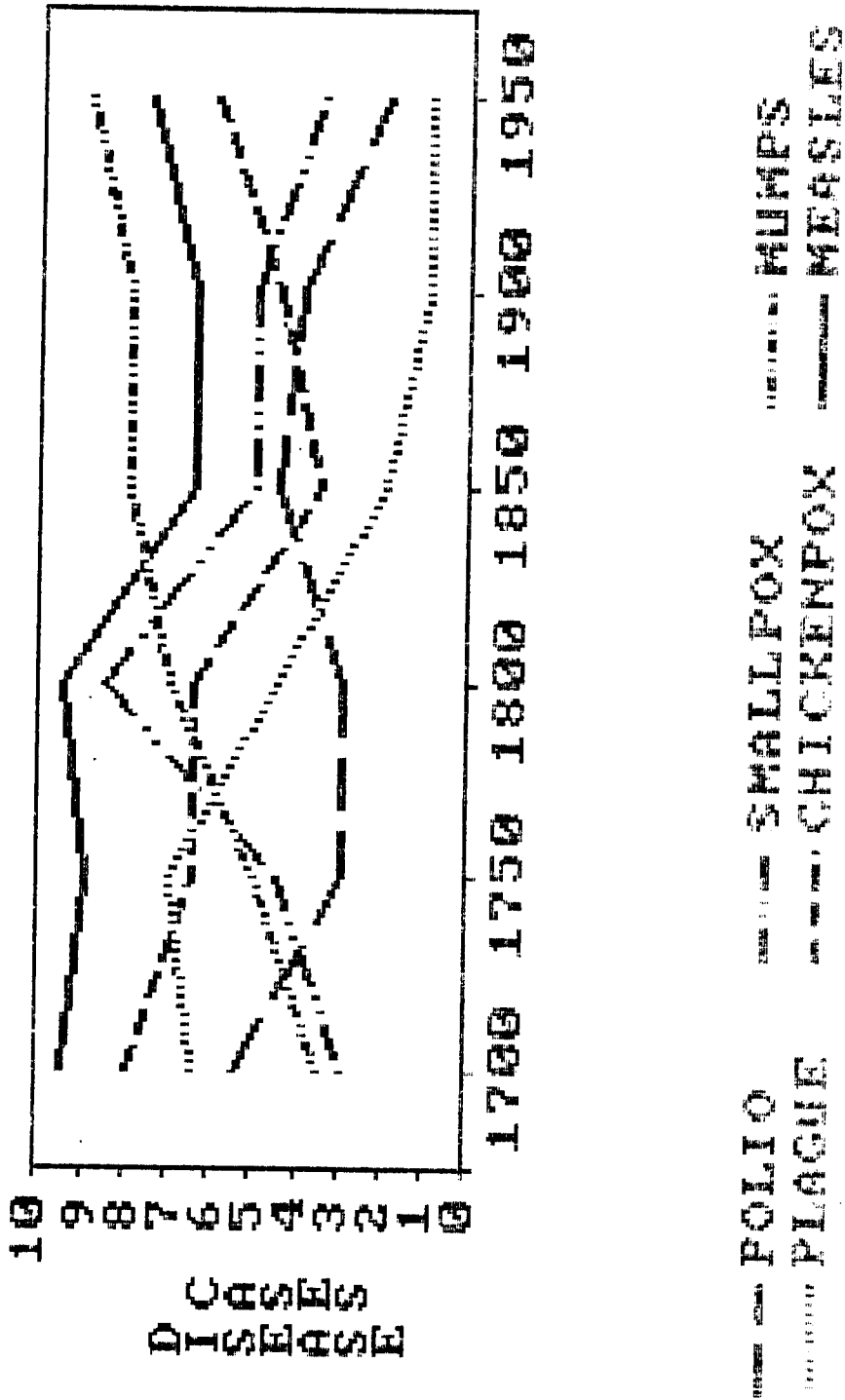
Low complexity, black-and-white, point-plot graph (Scale: 1 cm = 1.25 cm).



High complexity, black-and-white, point-plot graph (Scale: 1 cm = 1.25 cm).

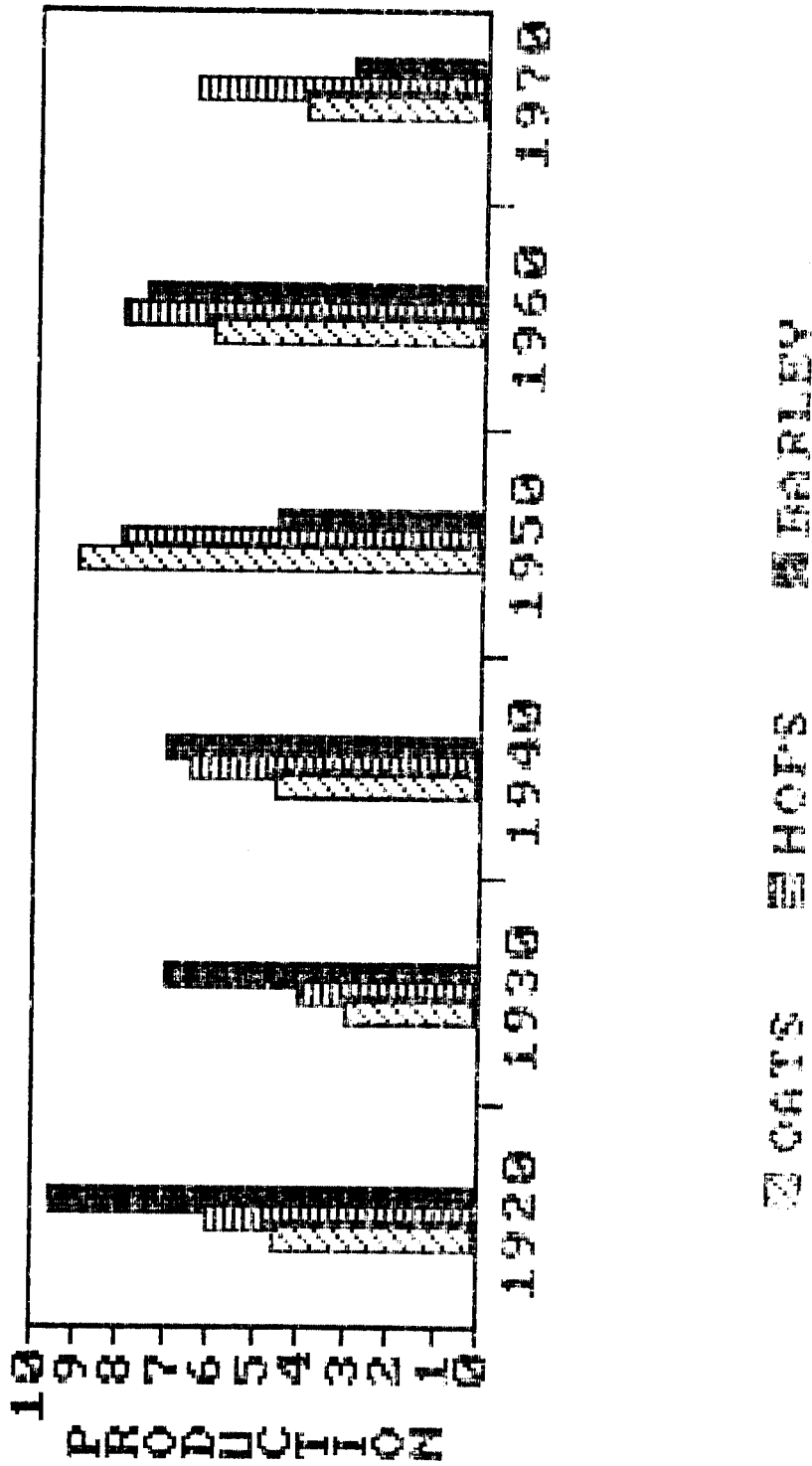


Low complexity, black-and-white, line graph  
 (Scale: 1 cm = 1.25 cm).

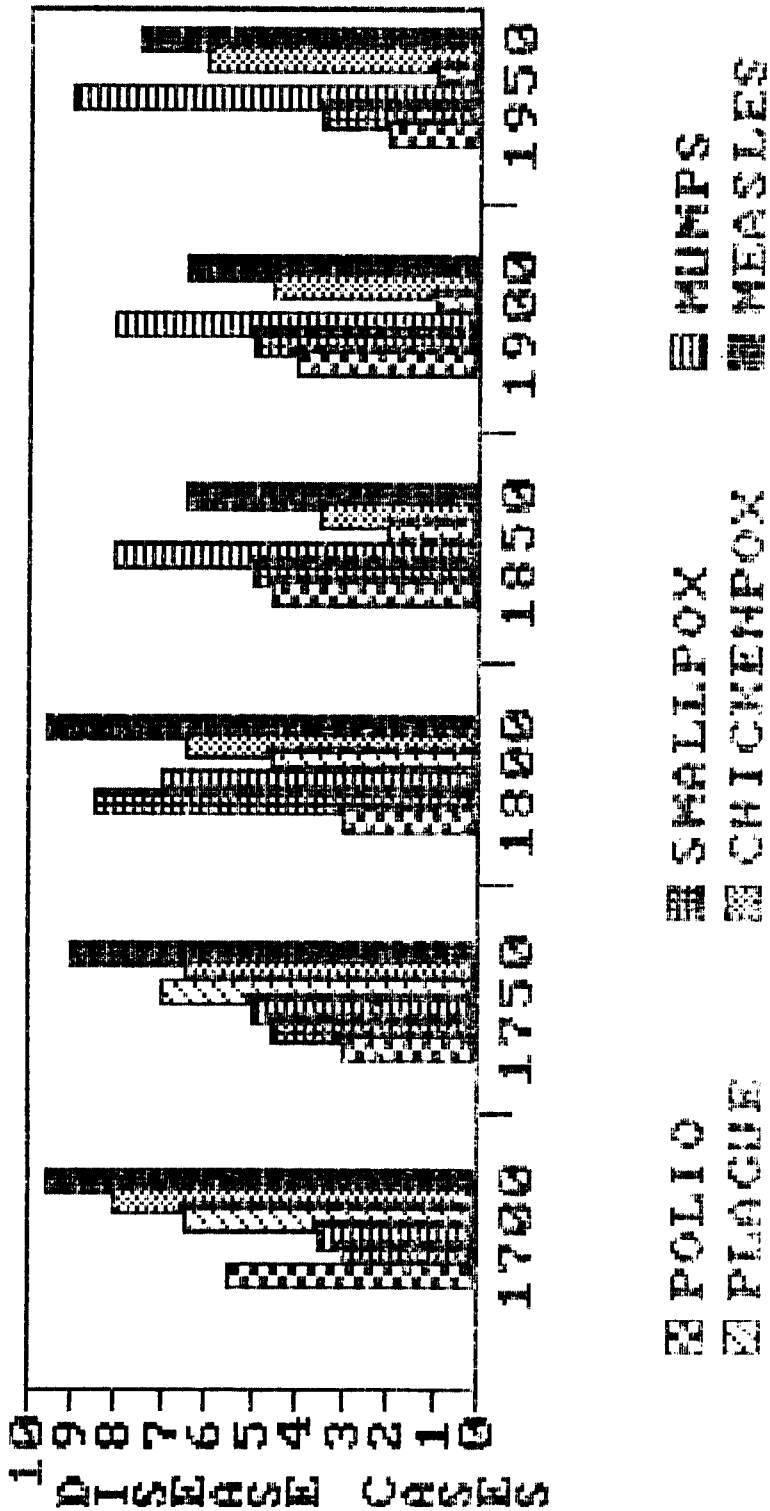


High complexity, black-and-white, line graph  
 (Scale: 1 cm = 1.25 cm).

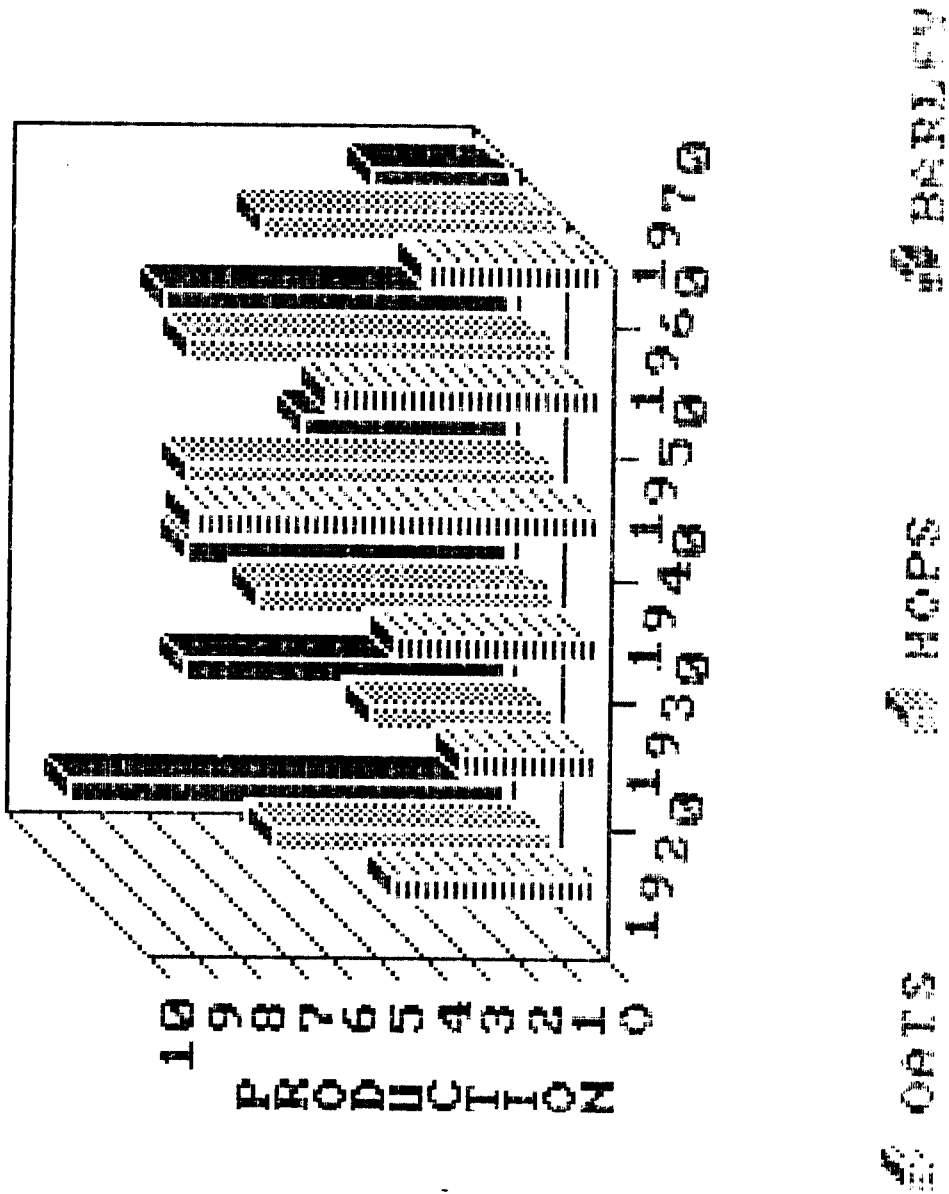




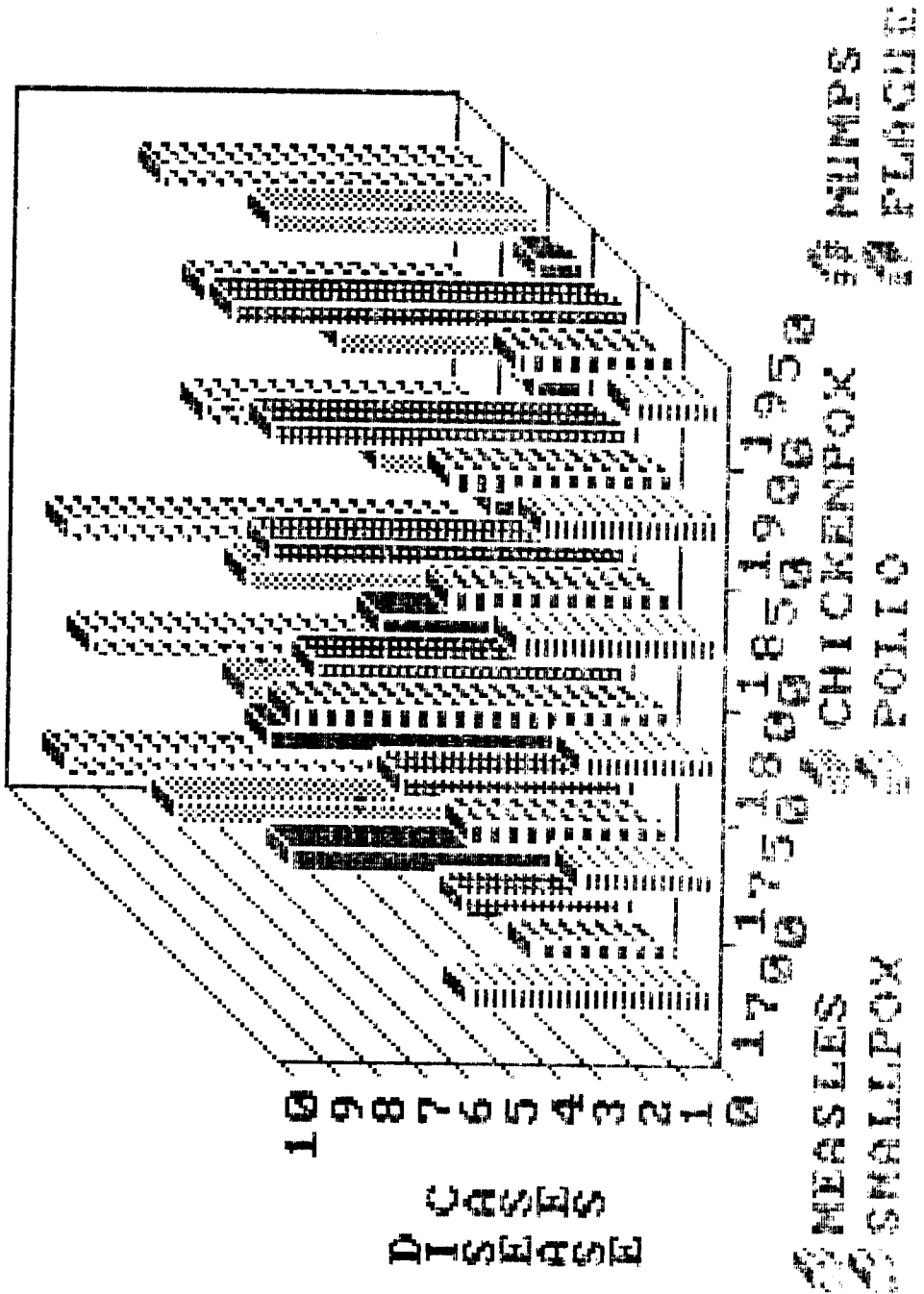
Low complexity, black-and-white, bar graph  
 (Scale: 1 cm = 1.25 cm).



High complexity, black-and-white, bar graph  
 (Scale: 1 cm = 1.25 cm).



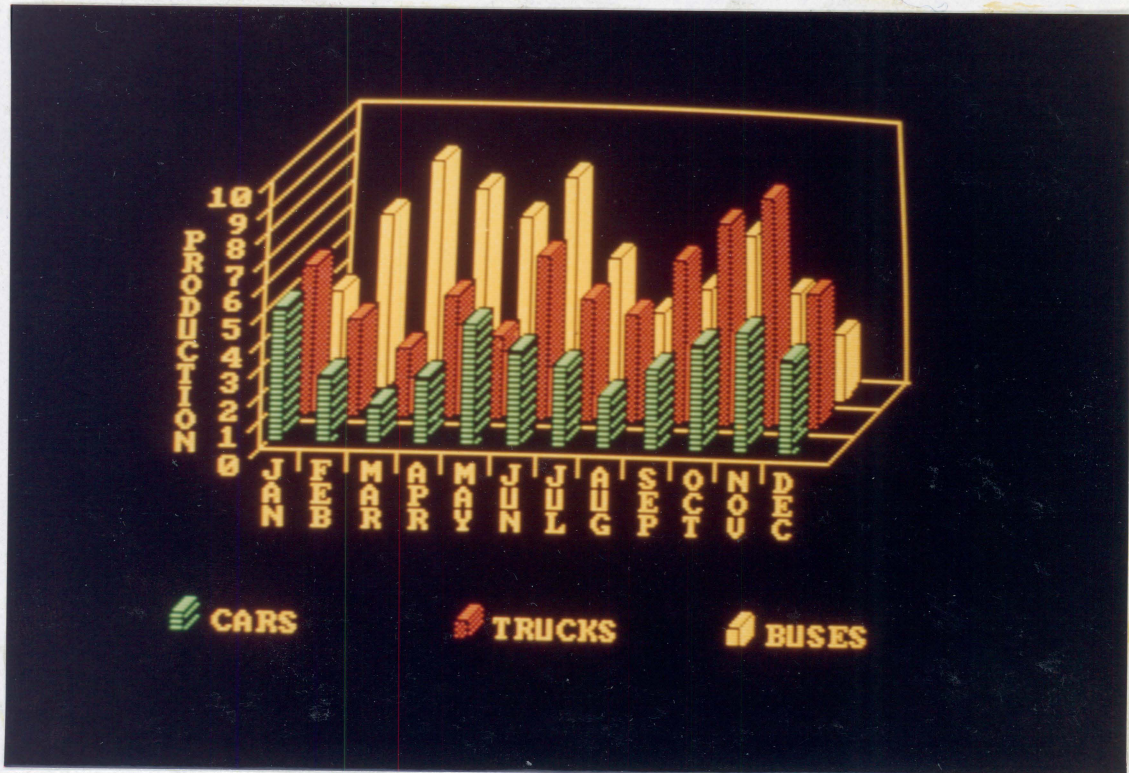
Low complexity, black-and-white, three-dimensional bar graph (Scale: 1 cm = 1.25 cm).



High complexity, black-and-white, three-dimensional bar graph (Scale: 1 cm = 1.25 cm).

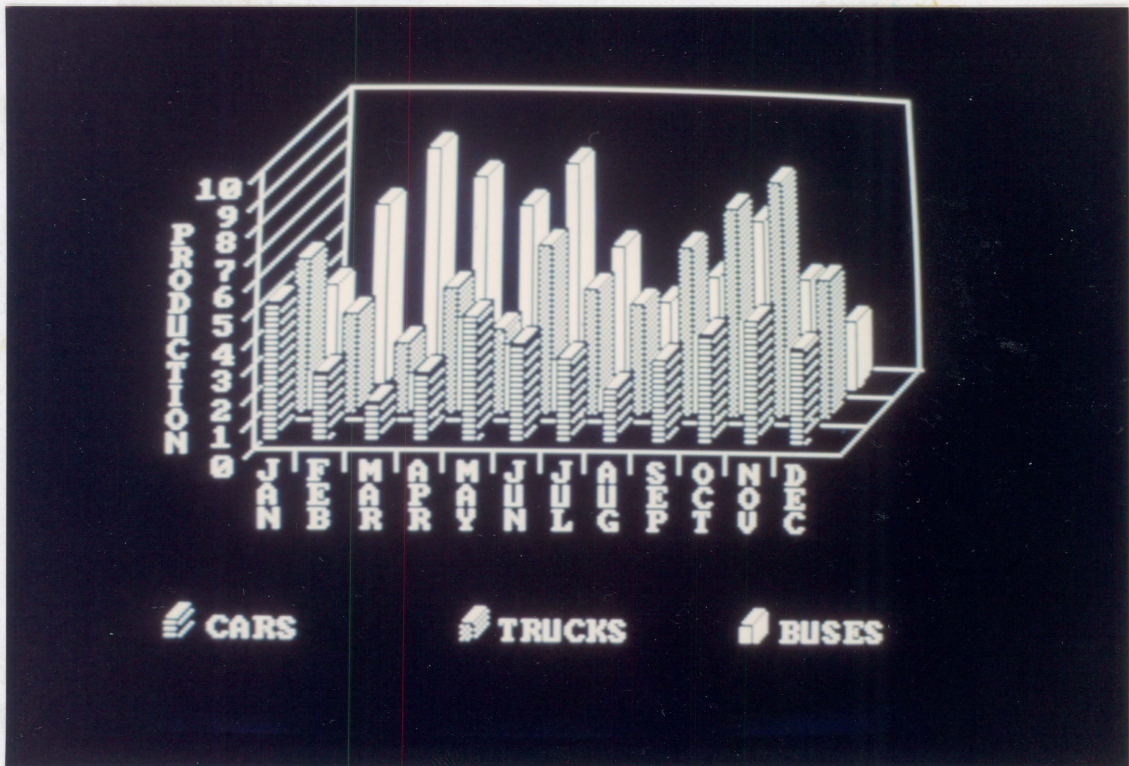
APPENDIX III

Examples of Color and Black-and-White Coding



Example of a color-coded graph  
 (Scale: 1 cm = 2.25 cm).





Example of a black-and-white coded graph  
(Scale: 1 cm = 2.25 cm).

APPENDIX IV

General Description of the Experiment for Subjects



## GENERAL INFORMATION ABOUT THE EXPERIMENT

The purpose of this experiment is to investigate the use of computer-generated graphics. The experiment will consist of one session lasting approximately 3.5 hours. A rest break of 5 minutes will be scheduled approximately every 45 minutes. An eye test is required before becoming eligible for participation. If you meet the vision requirements, you will then be given a training session consisting of 12 sample tasks. Next, you will begin the experimental session. This consists of tasks similar to those presented during the training period. At the end of the experiment you will be paid for your time. Your rate of pay will be \$3.00 per hour.

The data collected from your participation in this experiment will be treated in an anonymous fashion; that is, your name will not be connected with your data in any way. In addition, no part of this experiment will be painful or cause discomfort, however slight. If you feel uncomfortable, for any reason, you should STOP, and inform the experimenter immediately.

Please also note that many of the graphs that you will view, and the questions that you will be asked to answer, have been designed to be difficult. Your perceived score should not, in any way, be construed as reflecting any inherent lack of ability, learning disorders, or poor achievement.

After you have completed the session, please refrain from discussing the experiment with other individuals who might be future subjects until after September 15, 1984. Persons with prior knowledge of the experiment before participation might bias the experimental results.

This research is funded by a U.S. Department of Energy grant under the direction of the Management Systems Laboratory. The research team consists of Kenneth Gaylin, M.S. candidate, IEOR Department, and Dr. John G. Casali, Assistant Professor, IEOR Department. These individuals can be contacted at the address and phone number below:

Human Factors Laboratory  
Industrial Engineering and Operations Research  
Department  
Room 167 Whittemore Hall  
Virginia Polytechnic Institute and State  
University

Blacksburg, Virginia 24061  
(703) 961 5072

Please feel free to ask the experimenter any questions you might have. Answers to some questions, however, might have to be deferred until after the experiment has been completed, because we do not want to bias your responses.

APPENDIX V

Subjects' Informed Consent Document

## PARTICIPANT'S INFORMED CONSENT DOCUMENT

As a participant in this experiment, you have certain rights. The purpose of this sheet is to describe these rights to you and to obtain your written consent to participate.

1. You have the right to discontinue participating in the study at any time, for any reason. If you decide to terminate the experiment, inform a member of the research team and they will pay you for the portion of time you have participated.
2. You have the right to inspect your data and withdraw it from the experiment if you feel you should. In general, data are processed and analyzed after a subject has completed the experiment. At that time all identification information will be removed and all data will be treated with anonymity. Therefore, if you wish to withdraw your data, you must do so immediately after your participation is completed.
3. You have the right to be informed of the overall results of the experiment. If you wish to receive a synopsis of the results, include your address (three months hence) with your signature below. If after receiving the synopsis, you would then like further information, please contact the Human Factors Laboratory and a full report will be made available to you.

This research is funded by a U.S. Department of Energy grant under the direction of the Management Systems Laboratory. The research team consists of Kenneth Gaylin, M.S. candidate, IEOR Department, and Dr. John G. Casali, Assistant Professor, IEOR Department. These individuals can be contacted at the address and phone number below:

Human Factors Laboratory  
Room 167 Whittemore Hall  
Virginia Polytechnic Institute and State  
University  
Blacksburg, Virginia 24061  
(703) 961 5072

Further comments or questions can be addressed to Mr. Charles D. Waring, Chairman of the Institutional Review Board for the Use of Human Subjects in Research. He can be

contacted at the address and phone number listed below:

Office of Sponsored Programs  
301 Burruss Hall  
Virginia Polytechnic Institute and State  
University  
Blacksburg, Virginia 24061  
(703) 961 5283

The faculty and graduate student members of the research team sincerely appreciate your participation. They hope that you will find the experiment a pleasant and interesting experience. If you have any questions about the experiment or your rights as a participant, please do not hesitate to ask. We will do our best to answer them, subject only to the constraint that we do not want to pre-bias the experimental results.

Your signature below indicates that you have read and understand your rights as a participant (stated above), and that you consent to participate.

---

Signature

---

---

---

---

Printed name and address if you  
wish to receive a summary of  
the experimental results.

APPENDIX VI

Specific Instructions to Subjects

## SPECIFIC INSTRUCTIONS TO SUBJECTS

Please be seated and carefully read the following instructions. You may adjust your chair and location of the keyboard to positions which are most comfortable to you. You may re-adjust them at any point in the experiment. However, please do not attempt to adjust any controls on the television screens.

The format of the experiment is similar to that of a multiple choice quiz, however you will receive all necessary information and questions from the two television screens located to your left and center, and you will type your answers on the keyboard. The first thing you will see on the screen is:

PRESS THE SPACE BAR FOR THE NEXT QUESTION

This will appear on the left screen. When you press the space bar you will be presented with a question on the left screen. Please read the question and the multiple choice answers at this time. Some questions will not have multiple choices, but will instead ask you to type in a number. An example of each type of question is shown below:

1. How many cars were produced in January ?

Enter the correct NUMBER : \_\_\_\_\_

2. Which month has the highest death rate ?

F1. January

F2. July


F3. March

F4. August

Enter the correct FUNCTION KEY : \_

When the question provides multiple choice answers, the answer you select should be the best possible answer of those provided. After you have read, and become familiar with the question on the left screen, you may depress the space bar again, and a graph will appear on the center screen. You will need to use this graph to answer the displayed question. Please read and answer the question in as little time as possible while striving to be as accurate as you can in your answer. That is, do not rush yourself, but also do not use excessive amounts of time to

differentiate between minute (very small) differences in the available answers.





When you have selected an answer, press the key that is associated with that answer. Most answers will require the use of the function keys (F1, F2, F3,....F10) located at the leftmost side of the keyboard. During the experiment, only keys F1, F2, F3, F4, will be used. When the question asks you to enter a number, you should use the numbered keys on the topmost row of the keyboard. If you need to use a decimal point you should press the  key located at the bottom right of the keyboard, just above the space bar. All answers requiring a number to be entered should be accurate to one decimal place (one tenth of a unit represented on the scale of the graph). That is, answers requiring a number should have at least one numeral to the right of the decimal point. Some examples of correct and incorrect answers are listed below:

Correct answers

9.8  
0.7  
7.5  
3.0  
0.1

Incorrect answers

4  
2.08  
.003  
6.  
0.001

If you make a mistake, just re-type the function key, or backspace over the entered numbers using the  key at the middle right of the keyboard and re-enter them. Your answer will be highlighted in the answer space provided below the question on the left television screen. If you are satisfied that your answer is correct then press the  key which is next to the  key. Please be sure you have entered the function key or number that you desire before you press the  key, as you cannot go back and change an answer after depressing this key. You will then be prompted to rate the task you have just completed, using six preference scales that will appear on the left television screen. These ratings, and the procedures you should use to enter your rating selections are discussed in a separate handout that you will receive shortly. After you have completed all six ratings you will again be prompted with:

PRESS THE SPACE BAR FOR THE NEXT QUESTION

Each question will repeat this same format. When you are ready to continue again, you may press the space bar for the next question. Please note that you may rest for



several seconds or adjust your eyeglasses or chair during the time between answering the last question and beginning the next question, that is, during any time that the message to "depress the space bar for the next question" is present. If you feel you must speak to the experimenter you may do so at this time, however, it would be best if you waited for one of the longer 5 minute breaks that will be scheduled after every 45 minutes of testing. These longer breaks will be signaled by the message:

PLEASE TAKE A 5 MINUTE REST BREAK  
AND NOTIFY THE EXPERIMENTER THAT  
YOU ARE DONE WITH THIS SESSION

When this message appears you may get up, stretch, or leave the room for a 5 minute rest.

Although these procedures may seem confusing at first, they should come naturally to you after a short amount of practice. For this reason you will first receive a training period consisting of 12 questions. The training period will be similar to the experiment, except that the experimenter will be in the room to answer any questions you might have, and for the first 8 practice questions any incorrect answers will be signaled by an audible "beep". If an incorrect answer is given for these first 8 questions the computer will not go on to the next question, but will wait for you to try to get the correct answer. It will "beep" every time you put in the wrong answer during the practice period.

The last four practice questions will be presented in the exact same way as the experimental questions. That is, you will not hear a "beep" if you type in the wrong answer, and the computer will continue on to the next question even if the answer to the first question had been incorrect.

Remember, please read and answer the question in as little time as possible while striving to be as accurate as you can in your answer.

Please let the experimenter know when you are done reading the above instructions, and feel free to ask any questions you might have.

## RATING SCALE INSTRUCTIONS

Overview

After every task, you will be asked to give a rating on several descriptors associated with the task you have just completed. The descriptors are attentional demand, difficulty, task complexity, mental workload, stress level, and preference. Each of these descriptors will be defined on the left television screen so that you will maintain familiarity with the terms. Before you begin any experimental session, we will review:

1. A full set of the rating scales and the definitions of the terms used.
2. The steps you should follow in making your ratings on the scale.
3. How you should think about the ratings.

If you have any questions as these points are reviewed, please ask the experimenter. To familiarize you with this procedure we will give you a set of practice trials before you begin the experimental trials.

### Important Definitions

Before examining the rating sheets, I want to provide you with an important definition that you should keep in mind while performing the ratings. The word "task" appears on every rating sheet. "Task" refers to your task of answering each question that the computer displays on the left television screen, using the information on the graph displayed on the right television screen. That is, each question and graph constitutes an individual "task" which you answer and subsequently rate. The mechanics of entering the appropriate answer into the computer, use of the keyboard, or other problems such as glare on the television screens, or noise entering from outside the experimental room are not to be considered as factors when using the rating scales. Thus your ratings should be based only on the specific question and graph presented.

Now, let's proceed to the rating sheets (located on the last 6 pages of this handout). The first of them deals with Attentional Demand. As the rating scale indicates, attentional demand refers to the degree of concentration required (or the amount of attention required) to perform the task. Below the definition is a rating scale. After you have completed a task, you are to choose the one letter which best describes the attentional demand associated with the task. If the task required a great deal of concentration or attention, your rating should be to the right of center, that is, one of the letters G through K. If the task took little of your attention or concentration, your rating should be to the left of center.

The letters themselves have no implied meaning. They are simply used to designate equal intervals on a continuum of attentional demand from none to extremely high. The letters allow us to extract information from the rating scale with little likelihood of error.

We would like to rate your attentional demand in performing the task as accurately as possible. Please take as much time as necessary, and remember that as you move to the right on the scale, you are indicating higher attentional demand, and as you move to the left on the scale, you are indicating lower attentional demand. A rating of C represents less attentional demand than a rating of D.

The remaining five rating scales each contain a single descriptor definition and a single rating scale associated with that descriptor. Please read over the five remaining

definitions and examine their corresponding scales. The descriptions are not necessarily mutually exclusive (non-overlapping). Therefore, it is possible that your ratings on two or more scales may be similar. However, on each scale, please rate the descriptor based only on the definition given. If there are any questions, please feel free to ask.

### Rating Scale Steps

After you have completed an experimental task, the computer will display each full rating scale and its associated definition on the left television screen. Each scale will be presented successively, and in the same order. You should read the descriptor definition at the top of the screen, then carefully select a rating on the scale (displayed on the lower portion of the screen) which best describes the level of the descriptor you experienced. To be more specific, you should depress the appropriate letter on the keyboard that represents your choice or rating. This letter will then be highlighted in reverse video (black letters on a white background) to indicate that you have chosen it. If you would like to change your letter rating, just depress the key that corresponds to your desired rating. The computer will automatically highlight your new choice. If you are sure that you have chosen the appropriate rating, press the  key. You will then be given the next rating scale. Please note that you should be sure that you have entered the letter rating that you desire before you depress the  key, as you cannot go back and change an answer after depressing this key. When you have completed all of the ratings the next message you will see is :

PRESS THE SPACE BAR FOR THE NEXT QUESTION

When you are ready to go on to the next question, press the space bar.

Remember, please do not hurry in making your rating. Read the definition first, then carefully make the rating by entering the letter which is most appropriate.

### How You Should Think about the Rating Scales

Before making the ratings, there are several points that need to be emphasized. First, on all of your ratings you will be evaluating the graphs presented (in conjunction with the specific question asked) for a general user population, not yourself. You may assume that you are an

experienced member of that population. You should make the assumption that any problems you encounter are not problems you created. They should be considered as problems created by the graph presented in conjunction with the question that you are asked to answer. In other words, don't blame yourself; if the graph is deficient, blame the graph in your ratings.

Try to avoid the problems of nit picking an especially good graph and of saying that a graph is difficult to use if it is not difficult at all. These problems can result in similar ratings for graphs which are really quite different. Also, try not to over-react to small changes in a graph. This can result in ratings which are extremely different when the graphs themselves are really quite similar. Thus, to avoid any problems, just always "tell it like it is when making your ratings". If you have any questions, please ask the experimenter at this time.

ATTENTIONAL DEMAND refers to the portion of your total concentration required (or the amount of attention required) to perform the task.

YOUR RATING OF ATTENTIONAL DEMAND

A	B	C	D	E	F	G	H	I	J	K
NONE			MODERATE					EXTREMELY HIGH		

DIFFICULTY refers to how hard or difficult you found the task.

YOUR DIFFICULTY RATING

A	B	C	D	E	F	G	H	I	J	K
EXTREMELY EASY			MODERATE					EXTREMELY DIFFICULT		

TASK COMPLEXITY refers to how complicated or complex you found the task.

YOUR TASK COMPLEXITY RATING

A	B	C	D	E	F	G	H	I	J	K
EXTREMELY EASY			MODERATE					EXTREMELY COMPLEX		



MENTAL WORKLOAD is the integrated mental effort required to perform the task. It refers to the depth of thinking required by the task.

YOUR MENTAL WORKLOAD RATING

A	B	C	D	E	F	G	H	I	J	K
EXTREMELY LOW			MODERATE					EXTREMELY HIGH		

STRESS LEVEL refers to your emotional reaction while you performed the task. Stress may be considered your feeling of anxiety, concern, uneasiness and uncertainty brought on as a direct result of performing the task.

#### YOUR STRESS LEVEL RATING

A	B	C	D	E	F	G	H	I	J	K
ABSOLUTE CALM			MODERATE				EXTREME STRESS			

PREFERENCE LEVEL refers to how much you liked using the graph presented. In rating your preference level, you should consider how satisfied you were with the graph previously presented when you were using and looking at it in the task.

YOUR PREFERENCE LEVEL RATING

A	B	C	D	E	F	G	H	I	J	K
EXTREMELY DISLIKE				MODERATE				EXTREMELY LIKE		

APPENDIX VII

Subjective Mental Workload Rating Scales

ATTENTIONAL DEMAND refers to the portion of your total concentration required (or the amount of attention required) to perform the task.

YOUR RATING OF ATTENTIONAL DEMAND

A	B	C	D	E	F	G	H	I	J	K
NONE				MODERATE				EXTREMELY HIGH		

DIFFICULTY refers to how hard or difficult you found the task.

YOUR DIFFICULTY RATING

A	B	C	D	E	F	G	H	I	J	K
EXTREMELY EASY				MODERATE				EXTREMELY DIFFICULT		

TASK COMPLEXITY refers to how complicated or complex you found the task.

YOUR TASK COMPLEXITY RATING

A	B	C	D	E	F	G	H	I	J	K
EXTREMELY EASY			MODERATE					EXTREMELY COMPLEX		

MENTAL WORKLOAD is the integrated mental effort required to perform the task. It refers to the depth of thinking required by the task.

YOUR MENTAL WORKLOAD RATING

A	B	C	D	E	F	G	H	I	J	K
EXTREMELY LOW			MODERATE					EXTREMELY HIGH		



STRESS LEVEL refers to your emotional reaction while you performed the task. Stress may be considered your feeling of anxiety, concern, uneasiness and uncertainty brought on as a direct result of performing the task.

#### YOUR STRESS LEVEL RATING

A	B	C	D	E	F	G	H	I	J	K
ABSOLUTE CALM				MODERATE				EXTREME STRESS		

APPENDIX VIII  
Preference Rating Scale

PREFERENCE LEVEL refers to how much you liked using the graph presented. In rating your preference level, you should consider how satisfied you were with the graph previously presented when you were using and looking at it in the task.

YOUR PREFERENCE LEVEL RATING

A	B	C	D	E	F	G	H	I	J	K
EXTREMELY DISLIKE				MODERATE				EXTREMELY LIKE		

**The two page vita has been  
removed from the scanned  
document. Page 1 of 2**

**The two page vita has been  
removed from the scanned  
document. Page 2 of 2**