

Association for Information Systems AIS Electronic Library (AISeL)

ICIS 1994 Proceedings

International Conference on Information Systems
(ICIS)

12-31-1994

The Joint Effects of Interactions between Data Display and Task Variables on Task Performance

E. Vance Wilson

University of Colorado at Boulder

Theophilus Addo

National Sun Yat Sen University

T. Matsuzaki

IBM and Tokyo Institute of Technology

Margaret Tan

National University of Singapore

Follow this and additional works at: <http://aisel.aisnet.org/icis1994>

Recommended Citation

Wilson, E. Vance; Addo, Theophilus; Matsuzaki, T.; and Tan, Margaret, "The Joint Effects of Interactions between Data Display and Task Variables on Task Performance" (1994). *ICIS 1994 Proceedings*. 65.

<http://aisel.aisnet.org/icis1994/65>

This material is brought to you by the International Conference on Information Systems (ICIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICIS 1994 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

THE JOINT EFFECTS OF INTERACTIONS BETWEEN DATA DISPLAY AND TASK VARIABLES ON TASK PERFORMANCE

E. Vance Wilson

Graduate School of Business Administration and
Institute of Cognitive Science
University of Colorado at Boulder

Theophilus B. A. Addo

Information and Decision Systems Department
College of Business Administration
San Diego State University

ABSTRACT

Recent studies of data presentation focus on dependencies that exist between data displays and specific tasks. We propose that joint interactions between data display and key task variables contribute to these dependencies. This paper reports the findings of an experiment which investigates the effects of interactions between display format and two task variables: question type and question complexity. We find that both interactions significantly influence task performance.

1. INTRODUCTION

Powerful desktop computers and graphics software have empowered millions of computer users to design and produce their own data presentations. Since using inappropriate display formats to support a task is frequently inefficient (Jarvenpaa 1989), misleading (Addo 1994; Taylor and Anderson 1986), and unethical (Johnson, Rice, and Roemich 1980), it follows that the working productivity of these computer users will be increased if research can guide them to select the best display format for each task.

It is now generally accepted that performance outcomes of using a given data display format are dependent on the task which is to be accomplished (Tan and Benbasat 1990). In factorial research designs, these dependencies have been observed as interactions for which the causes are unclear. Dickson, Senn, and Chervany state that "the existing research has clearly demonstrated that there is an important system/user/decision interaction operating" (1977, p. 921). Zmud, Blocher, and Moffie report findings that "strongly support the existence of an interaction effect between report format and task complexity [and] suggest that some (still unknown) learning phenomenon may exist regarding report formats, task complexity, and decision behavior" (1983,

p. 190). It is important that research follow up these observations, with the goal of understanding what factors cause display-task dependencies.

In this paper, we report the results of an experiment that investigates the joint effects of interactions between display format and two task variables: question complexity and question type. The paper has two related objectives. First, the findings may help both to explain inconsistencies that have marked past research and to provide guidance for future researchers, as well as data presentation designers and users. Second, the findings may support the long-term goal of developing grounded data presentation theory.

In the next section, background literature is reviewed. Research hypotheses are then proposed, and a research methodology is outlined. This is followed by presentation of results and discussion.

2. INTERACTIONS BETWEEN DISPLAY AND TASK VARIABLES

In her frequently-cited survey of data presentation studies, DeSanctis (1984) describes a framework for research in the

field. We have adapted this framework in order to position our study within the research stream (see Figure 1). The framework proposes three key inputs that influence performance outcomes of using presentations: (1) individual characteristics; (2) display characteristics; and (3) task characteristics. These inputs are considered to have main effects and, potentially, interactive effects on human cognitive processes involved in acquiring and using information from displays.

Our focus within the framework is on the relationship between a display characteristic (display format) and two task characteristics (question type and question complexity). Display format is an important variable for study in the context of computer presentations, as computer users typically have much more control over the display format than they have over other factors, such as the data that is to be displayed or the task that needs to be accomplished. Since users can readily change their display format criteria, findings of research involving display format have the potential to be of practical, as well as theoretical, importance.

The relationship of task characteristics to display format is also important. Question type and question complexity are two key dimensions of question tasks, i.e., tasks that focus on processes of information acquisition rather than decision making or problem solving. In the following sections, we review research which suggests that question type and question complexity variables jointly interact with display format.

2.1 Display Format by Question Type Interaction

The cognitive fit theory developed by Vessey (1991) predicts display-task dependencies resulting from interaction between display format and the *type of question* that is asked. Cognitive fit is defined as “a cost benefit characteristic that suggests that, *for most effective and efficient problem solving to occur, the problem representation and any tools or aids employed should all support the strategies (methods or processes) required to perform that task*” (Vessey and Galletta 1991, p. 64, emphasis in original).

Vessey describes two fundamental types of question tasks: *symbolic* questions, which require precise values, and *spatial* questions, which require comparisons. She correspondingly proposes that tables are symbolic representations of numeric data and graphs are spatial representations. Cognitive fit and resulting superior task performance are held to occur when display format and question type support formation of a uniform mental representation of the

problem. This is predicted to occur when symbolic representations (tables) are matched to symbolic questions and spatial representations (graphs) are matched to spatial questions.

[A] spatial representation does not have to be used to solve a spatial task; nor does a symbolic representation have to be used to solve a symbolic task. For example, a problem solver might determine a trend from a table or extract a specific numeric value from a graph. When the information in the problem representation and the task do not match, however, similar processes cannot be used to both act on the problem representation and solve the problem, and the mental representation will have to be transformed. [Vessey and Galletta 1991, pp. 68-69]

Initial empirical testing of the cognitive fit theory found support for the postulated interaction between display format and question type on a response time measure, but only partial support on accuracy (Vessey and Galletta 1991). Contrary to predictions, tables were more accurate than graphs for answering spatial questions. A contemporaneous study conducted by Coll (1992) measured accuracy using *retrieve specific values* and *retrieve relative information* question types that are closely similar to cognitive fit theory's symbolic and spatial questions. Coll's findings completely support the interaction predicted by cognitive fit theory, namely, tables were superior in accuracy performance for the retrieve specific values questions and graphs were superior for the retrieve relative information questions. In their discussion, Vessey and Galletta write that their spatial questions possibly were “too simple for spatial representations to have an advantage over tables” (1991, p. 79), suggesting that joint analysis with question complexity may be important in applying the cognitive fit theory to question tasks. Vessey and Galletta recommend that further research be conducted to address this issue.

2.2 Display Format by Question Complexity Interaction

Question complexity is defined herein as the component of difficulty in performing question tasks attributable to the behavioral requirements of the task, rather than to characteristics of the individual performing the task. This definition is consistent with larger task definitions proposed by Hackman (1969) and Wood (1986). Several studies have found that question complexity directly affects question task performance (e.g., Davis 1986; Addo 1989). The interaction of question complexity and display format has not been directly explored, but two studies using related

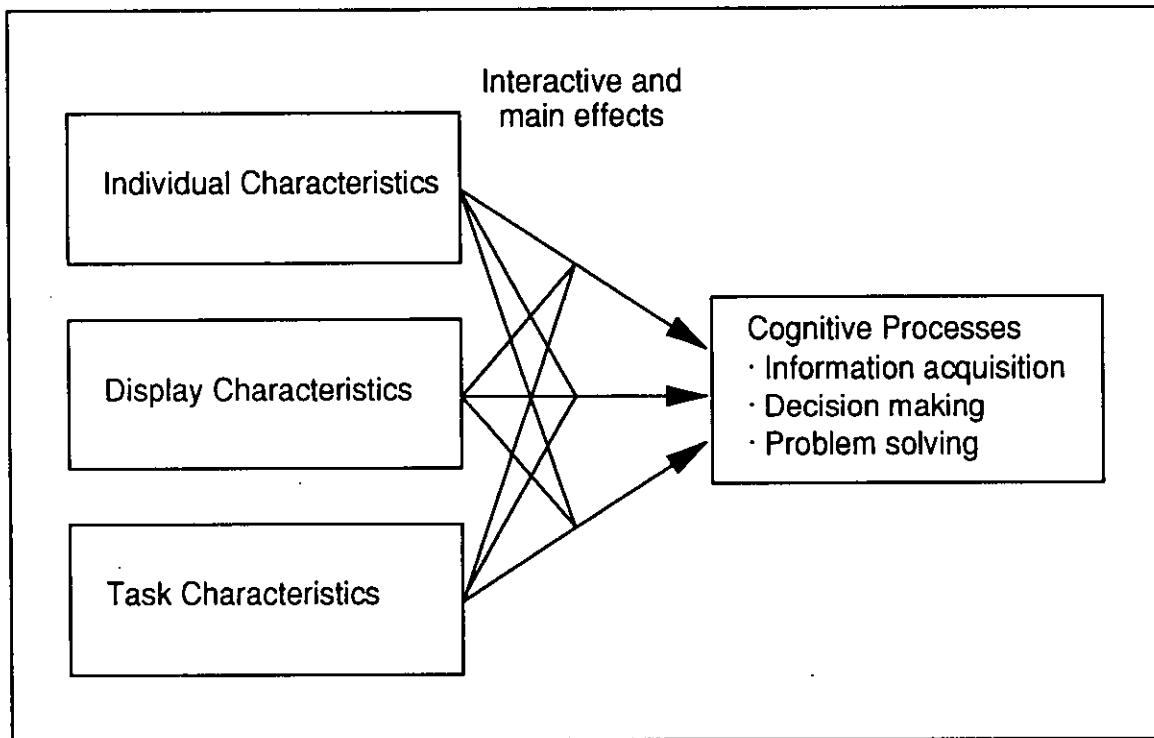


Figure 1. Framework for Presentation Research
(adapted from DeSanctis 1984)

variables have reported interactions. Schwartz and Howell (1985) employed a hurricane-tracking scenario which varies time constraint. Time constraint has been found to contribute to general task complexity (Wright 1974), suggesting that time constraint is equivalent to increased question complexity. Schwartz and Howell report that subjects under time constraint reach better final decisions using graphs than tables. No difference was found under test conditions without time constraint.

In a study that employed a production scheduling problem, Remus (1987) tested subjects under low and intermediate levels of environmental complexity created by using data sets with different levels of variance. Remus finds that decision performance is better with tables in low complexity environments and with graphs in intermediate complexity environments. Environmental complexity, as operationalized by Remus, has aspects of information complexity as well as question complexity. Information complexity describes the component of difficulty in performing question tasks attributable to the data display, rather than to the task behavior requirements or the characteristics of the

individual performing the task. This construct was explicated by Bertin (1983) and was subsequently operationalized in a series of metrics developed by Yoo (1985), Lauer (1986), and Joyner (1989). A primary effect of increased variance is to make the question more difficult to answer, thus creating more stringent behavior requirements regardless of characteristics that individuals may bring to the task. Therefore, Remus' findings are potentially relevant to question complexity research.

These studies suggest that question complexity may interact with display format. However, this inference is made from tests of related variables rather than question complexity. Further research should extend these findings using a question complexity variable.

2.2 Summary

Characteristics of users, displays, and tasks all potentially influence the processes involved in acquiring information from data displays. Because unidentified dependencies are

acknowledged to exist between display and task, these areas are especially important for study. We propose to study interactions that may exist for task performance measures between key characteristics of display (display format) and task (question complexity and question type). Research suggests that these interactions are appropriate topics for further study.

3. HYPOTHESES

Hypothesis 1 implements a test of the cognitive fit theory's predictions. It replicates findings by Vessey and Galletta and by Coll that task performance is better under conditions where display format (DF) and question type (QT) are matched for their cognitive characteristics. We state the following hypothesis:

H1: A significant DF x QT interaction will be found where tables produce superior performance outcomes for symbolic question types and graphs for spatial question types.

Hypothesis 2 investigates display format and question complexity (DF x QC) in extension of findings by Schwartz and Howell and by Remus that variables similar to these have interactive effects on task performance. We state the following hypothesis:

H2: A significant DF x QC interaction will be found where tables produce superior performance outcomes at low levels of question complexity and graphs at high levels of question complexity.

Beyond replicating and extending the antecedent studies, findings which support both the hypotheses will indicate that joint interaction of both sets of factors must be considered in the design of future research. Such findings will support admonitions that already have been made against conducting simplistic, single-factor studies in this area (e.g., Coll 1992), i.e., main effects cannot be unambiguously interpreted where the variables interact. More importantly, further implications may be shown to exist for factorial research designs.

Results from two of the reviewed studies provide an exercise for these implications. Vessey and Galletta and Coll investigated the effects of interaction between display format and question type, using factorial experiments. In each study, numerous variables, including question complexity, were controlled at a single level, as is common in experimental designs. However, use of a single-level

control strategy assumes either that the investigated experimental effects are generalizable across levels of the controlled variables or that the findings will be interpreted only for the level at which the variable is controlled.

These assumptions raise problems in interpreting findings between the studies in the event that display format proves to interact with question complexity as well as question type. The existence of a DF x QC interaction will suggest that findings of the two studies apply narrowly to the level of question complexity that each study operationalized. Since there is no indication that question complexity was controlled at the *same* level in these studies, the findings may not be equivalent, despite their logical similarities.

This lack of equivalence could figure in contradictory findings that occurred between the studies in accuracy performance for spatial question types. Vessey and Galletta's subjects answered these questions more accurately using tables ($p < .0005$), but Coll's answered more accurately using graphs ($p < .0001$). Does this contradiction arise from chance, from differences in measurement precision in the studies, or from a spurious source, such as differential question complexity?

Our study addresses this issue by investigating the joint effects of display format interactions with question type and question complexity. It is highly unlikely that these specific interactions are the sole cause of contradictions and equivocality among data presentation studies, but we anticipate that investigation of their joint effects will add incrementally to the existing research by expanding the understanding of factors responsible for display-task dependencies. In this manner, our study potentially contributes to the field by enhancing the *post hoc* interpretability of the existing literature and by guiding the direction of future research.

4. RESEARCH METHODOLOGY

4.1 Experimental Design

A laboratory experiment employing a repeated measures, subjects-by-treatments design was used in this study. Independent variables — display format by question type by question complexity — were tested in a balanced $2 \times 2 \times 2$ factorial design for their effects on task performance accuracy and response time dependent variables. Each subject received all eight experimental treatments. Carry-over, latency, and learning effects were controlled by randomization of treatment administration order.

Table 1. Operationalization of Experimental Questions

Question Type	Question Complexity	Sample Question Text
Symbolic	Low	"What is the quantity in period 2?"
	High	"What is the combined quantity in periods 1 and 7?"
Spatial	Low	"Of periods 1 and 2, which has the larger quantity?"
	High	"Of periods 1, 3, 4, and 7, which period has the quantity closest to the midpoint between the highest and lowest?"

4.2 Independent Variables

Display format. Two levels of display format were used as independent variables in this study: tabular and graphical. Previous researchers have used a variety of graphical formats; in this study, we follow Coll in using vertical bar graphs. Display formats conformed to published design guidelines (Ives 1982).

Question type. Question type development was guided by cognitive fit theory (Vessey 1991), which postulates two levels of question type: symbolic and spatial. Each question type was represented by a single question at each level of question complexity (see Table 1).

Question complexity. Question complexity was operationalized using guidelines from a metric developed by Addo (1989). In developing this metric, Addo demonstrated that question complexity is a function of three factors: a *step* factor, an *image isolation* factor, and a *polarity* factor. The mathematical formulation of this metric is as follows:

$$Q_c = T_s + T_i + T_p$$

where: Q_c = question complexity;
 T_s = time value attributed to step factor;
 T_i = time value attributed to image factor; and
 T_p = time value attributed to polarity factor.

The step factor refers to the number, as well as type(s), of steps that must be performed in order to acquire information from a graph. These steps, in increasing order of difficulty, are identification, scan, comparison, and estimation. The greater the number and difficulty of the required steps, the greater the contribution to complexity.

Image isolation refers to the effort required to isolate the relevant image from the rest of a data display in order to

obtain the correct answer to a question. An important aspect of this factor deals with the relative logical proximity of the components to which a question directs the user. The higher the degree of logical proximity of these components, the easier it is to isolate the relevant image. Consider a time-series line or bar graph showing profits for three companies over twelve time periods. Using this graph, the correct answer to the question "*In periods 1 and 2, which company shows the largest difference in profits?*" will be obtained in a shorter period of time than would be required to obtain the correct answer to a similar question involving periods 1 and 8. The only difference between these two questions is in the relative proximity of the relevant time periods specified. The phenomenon of proximity has been noted by several researchers, within various contexts, as an important aspect of graphics use and research (see the proximity compatibility hypothesis in Carswell 1992; as well as Bertin 1983; Vickers 1979; Leeuwenberg 1968). Addo found this factor to be the most significant determinant of question complexity.

Polarity refers to the positive or negative phrasing of a question, for example, "*Which is longer?*" versus "*Which is shorter?*" Even though question polarity has been shown to affect performance (Parkman 1971), Addo found only partial support for this phenomenon.

Addo's metric guided the development of the set of questions shown in Table 1. In each low-complexity/high-complexity pair of questions, the low-complexity version requires the performance of fewer steps (i.e., identifications, scans, comparisons, and estimations) than the high-complexity version. Additionally, each low-complexity question requires the acquisition of information (i.e., isolation of relevant images) from point(s) that are in closer logical proximity of each other than those in the corresponding high-complexity version. The present study does not use polarity in its determination of question complexity; all questions are positively phrased.

4.3 Controlled Variables

Color. Color was controlled at optimal levels for each display format, as recommended by Hoadley (1990). Tables were presented as black text on a white background; bar graphs were presented as red bars with black outline on a white background.

Information complexity. Complexity of information presented via the displays was controlled at a low level, as determined by a metric developed for this purpose (Joyner 1989).

4.4 Dependent Variables

Accuracy and response time were chosen as dependent variables. These variables are frequently used in information processing research (Pachella 1974), and it has been recommended that they be jointly measured in data presentation research (Jarvenpaa 1989).

Accuracy. Accuracy of task performance was measured as a correct/incorrect dichotomy for all question treatments. Although symbolic questions are amenable to quantification, spatial questions involve a nonquantitative comparison. Therefore, accuracy measures for both question types were converted to binary values in order to support an equivalent data format. An accurate response for symbolic questions was defined to be a numeric answer within $\pm 5\%$ of the precise answer and for spatial questions as the correct selection from alternative choices.

Response time. Response time of task performance was measured as the time in fractions of seconds that elapsed between the initial display of the experimental treatment and the subject's completion of the experimental task, denoted by pressing the return key.

4.5 Subjects

The experiment was first administered to seven pilot test subjects, whose results were not included in the final data, and then to 36 experimental subjects. The subjects were volunteers drawn from a population of senior and graduate business students attending San Diego State University. Mean length of full-time employment experience for the sample group was over eight years, and management experience averaged over three years.

4.6 Procedure

The experiment was administered using a Macintosh IIfx computer with a 13 inch Apple RGB monitor. Although

implemented on a Macintosh computer, the software application was designed for generic appearance and operation. No mouse was attached to the computer during the experiment, and no pull-down menus were used. The software application was self-contained; it explained the experiment to the subjects, provided sample treatments for practice purposes, administered the actual experiment, and collected response time and accuracy performance data.

Each subject was welcomed by the researcher and seated at the test computer. After obtaining the subject's written consent to participate in the experiment, the researcher left the room. Only the participating subject was present in the room during the experiment. Subjects progressed sequentially through the software application by entering data as requested and by pressing the return key to move to the next screen (see sample screens in Figure 2). The application provided a tutorial and a training set of displays which provided the subject with feedback on the accuracy of his or her responses. Prior to beginning the experimental treatments, each subject was instructed by the software to "answer each question as accurately and quickly as possible." All subjects successfully completed the experiment, with an average completion time of approximately 30 minutes.

5. DATA ANALYSIS AND RESULTS

The research hypotheses investigate whether interactions occur between display format and two task variables: question type and question complexity. Descriptive statistics for the collected data are shown in Table 2. The accuracy and response time data were not correlated for any of the test conditions ($p > .05$), indicating that no significant speed-accuracy trade-off occurred within conditions in the experiment.

Repeated measures ANOVA was chosen to analyze the effects of these variables on accuracy and response time measures of task performance. This method offers the desirable features of analyzing interactions and handling dependent data (data that is obtained through repeated measures of the same subject). However, several violations of the assumptions for repeated measures ANOVA were encountered in the data. First, accuracy data were encoded to produce an equivalent binary format for all task conditions. ANOVA is not typically employed for analysis of dependent binary data, but Monte Carlo studies conducted by Mandeville (1969, 1972) on one- and two-factor repeated measures ANOVA designs indicate that use of dependent binary data has only moderate effects on both critical α and power when sample sizes are greater than 30. He states that "it appears safe to suggest that repeated measures analysis of variance may be recommended to a researcher who wishes to retain the [binary] item response

7 of 36

Total quantity is 26731

What is the quantity in period 3?

Period	Quantity
1	4759
2	932
3	3975
4	2881
5	1370
6	1290
7	171
8	3629
9	989
10	698
11	2737
12	3300

Please enter a number answer
and then press the RETURN key.

12 of 36

Of periods 8 and 9, which has the larger quantity?

Please enter a number answer
and then press the RETURN key.

**Figure 2. Sample Screens from Test Administration
Showing Tabular and Graphical Display Formats**

Table 2. Descriptive Statistics of Accuracy and Response Time Measures (n = 36)

Question Type	Question Complexity	Display Format	Accuracy* Mean (Std. Dev.)	Response Time** Mean (Std. Dev.)
Symbolic	Low	Table	1.00 (.00)	359 (124)
		Graph	.75 (.44)	614 (299)
	High	Table	.86 (.35)	1477 (831)
		Graph	.86 (.35)	1240 (669)
Spatial	Low	Table	.97 (.17)	409 (184)
		Graph	1.00 (.00)	392 (199)
	High	Table	.58 (.50)	2196 (1169)
		Graph	.58 (.50)	1551 (815)

* Percent correct

** Ticks (60 per second)

Table 3. Repeated Measures ANOVA Results on Accuracy and Response Time (n = 36)

Source of Variation	DF	SS	Accuracy		Response Time		
			F	Sig.	SS	F	Sig.
Display Format (DF)	1	.22	2.20	.147	1,859,074	6.06	.019
Question Type (QT)	1	.50	5.38	.026	3,315,097	8.90	.004
Question Complexity (QC)	1	3.13	26.52	.000	99,004,728	172.62	.000
DF x QT	1	.35	2.76	.106	2,087,435	8.25	.007
DF x QC	1	.22	1.93	.173	5,662,453	18.58	.000
DF x QT x QC	1	.35	2.48	.124	83,742	.38	.540

as the basic datum in his analysis" (1972, p. 321). This recommendation supports our use of repeated measures ANOVA for analysis of binary accuracy data, but correspondingly suggests that we interpret with caution any significance findings that are near to the critical α probability level.

Second, significant skewness and heterogeneity were observed in the response time data. However, in such cases as this study, where the design is balanced and independent variables are implemented at only two levels (producing equivalence between univariate and multivariate tests of significance), repeated measures ANOVA has been found

to be robust to violation of multivariate normality and homogeneity of variance assumptions (O'Brien and Kaiser 1985). Therefore, the data were analyzed without transformation. Effects were tested using the following ANOVA model for each dependent measure:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \pi_l + \alpha\beta_{ij} + \alpha\gamma_{ik} + \alpha\beta\gamma_{ijk} + \epsilon_{ijkl}$$

where: μ = a constant;
 α_i = display format ($i = 1,2$);
 β_j = question complexity ($j = 1,2$);
 γ_k = question type ($k = 1,2$);
 π_l = experiment subject ($l = 1$ to 36); and
 ϵ_{ijkl} = error.

Results of the analysis are shown in Table 3.

5.1 DF x QT and DF x QC Interactions

Substantially greater overall effects were found in response time measures than in accuracy measures. As previously discussed, we anticipated that the accuracy analysis in this study could show moderately reduced power due to the binary format of the accuracy data (Mandeville 1969, 1972). However, the F -ratios observed for accuracy in analysis of both main effects and interactivity are much smaller in relation to those of response time than would be expected from this single cause. Therefore, we interpret the results to indicate that our subjects set relatively inflexible accuracy standards and used the amount of time necessary to meet those standards, causing the treatment effects to be seen primarily in the response time measure, as was previously reported by Bettman and Zins (1979) and Jarvenpaa (1989). It is known that changing the payoff structure which subjects work under results in adjustment of their priorities for accuracy versus speed (Pachella 1974). It follows that effects seen in the present study for response time may be anticipated to trade off toward greater accuracy effects under conditions that reward speedy performance or constrain the amount of time which subjects are given to complete the tasks (Bettman and Zins 1979).

On response time measures, an interaction was found between display format and question type (DF x QT), supporting Hypothesis 1 ($F[1,35] = 8.25, p = .007, \omega^2 = .164$). The findings for Hypothesis 1 support predictions of cognitive fit theory and generally replicate the display format by question type interactions reported by Vessey and Galletta and by Coll. The index, ω^2 , shows the proportion of variation in a dependent variable accounted for by the effect (Hays 1963). Our results indicate that the DF x QT interaction accounts for 16.4% of the variation in response time.

Interaction was also found between display format and question complexity (DF x QC), supporting Hypothesis 2 ($F[1,35] = 18.58, p < .0005, \omega^2 = .322$). These results are similar to the interactions between display format and task-related complexity variables observed by Schwartz and Howell and by Remus. The DF x QC interaction accounts for 32.2% of the variation in response time.

Three-way interaction, DF x QT x QC, on the response time measure was not significant ($F[1,35] = .38, p = .54$). Graphs of the observed interactions are shown in Figure 3. Interactions on the accuracy measure are not significant, but are included in Figure 3 for completeness.

6. DISCUSSION AND CONCLUSIONS

Two major conclusions may be drawn from the findings of this research. First, the investigated interactions proved to be significant. Effects of the interactions were also moderately strong, accounting for approximately 16% (DF x QT) and 32% (DF x QC) of response time variation. The confounding effects that this level of interactivity can have on simplistic studies cannot be overstated. Indeed, one subset of the findings of this study readily supports the proposition that tables are better than graphs — i.e., the accuracy performance of low complexity symbolic tasks ($t = 3.42, 35$ df, $p = .002, \delta = .57$). Another subset supports the proposition that graphs are better than tables — i.e., the response time performance of high complexity spatial tasks ($t = 3.55, 35$ df, $p = .001, \delta = .57$). It is clear that neither of these conclusions is generalizable; the simple effects cannot be meaningfully interpreted due to interactivity among the variables.

Even tests of two-way interaction fail to capture important dimensions of the phenomena. Neither of the separate interactions — display format with question type or display format with question complexity — adequately describe the *joint effects* of these interactions on task performance. This finding suggests that the theory of cognitive fit, which considers only the type of task, will benefit from further theoretical development which addresses task complexity characteristics as well.

Second, the results support an expanded structural view of the nature of the dependencies that exist between display format and task characteristics. Specifically, a pattern of relative performance emerges from the findings which addresses the old question, "*Which are better, tables or graphs?*" (see Figure 4). In the condition where tasks are simplistic and symbolic in nature, tables are clearly superior to graphs. Conversely, where tasks are complex and spatial in nature, the scale of performance tips in favor

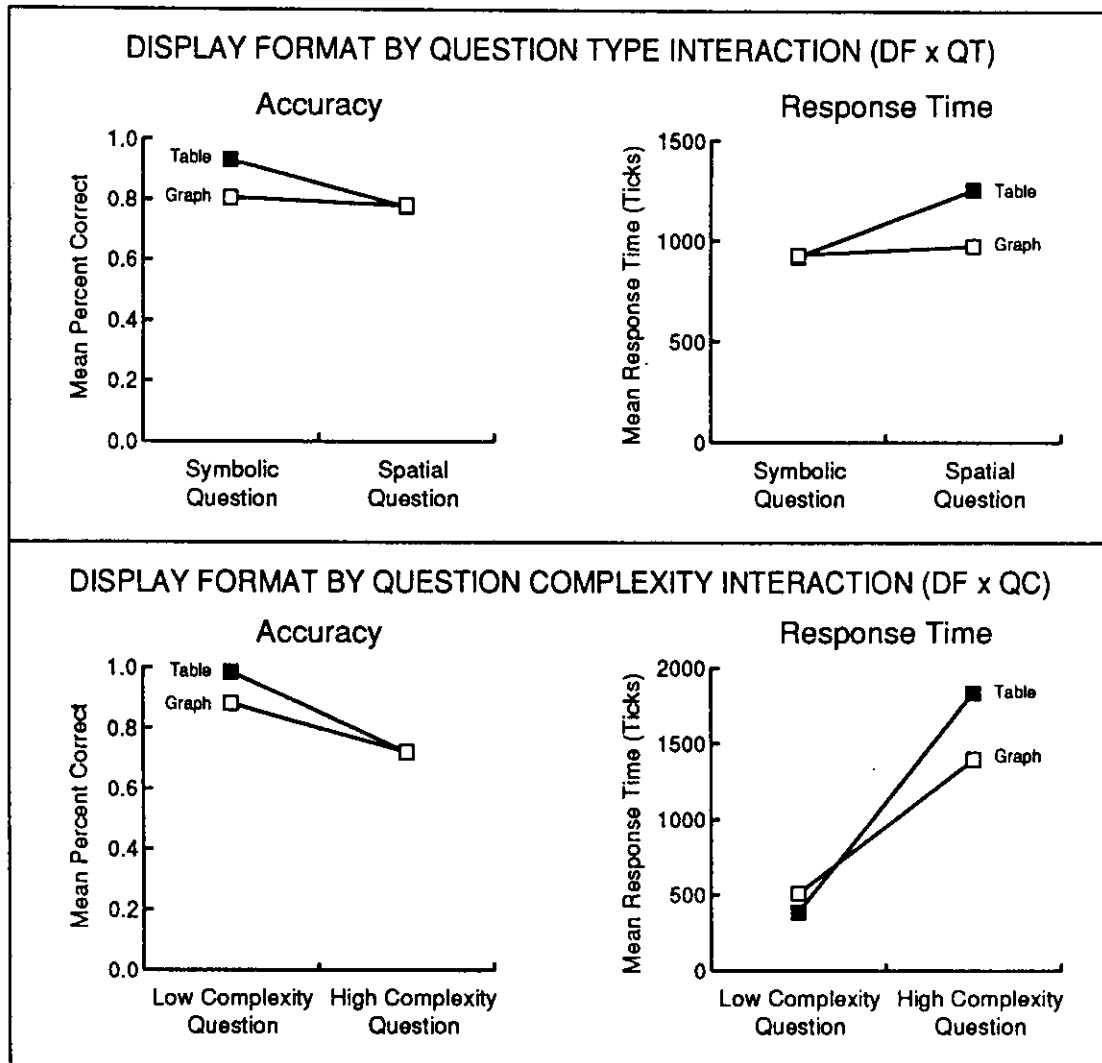


Figure 3. Interaction Graphs

of graphs, especially in terms of response time. Over some intermediate area, performance differences between tables and graphs are minimal.

These findings may be applied to *post hoc* interpretation of data presentation studies. We previously described the contradictory results for accuracy performance reported in the Vessey and Galletta and Coll studies, which investigated interaction between display format and question type. Our findings clearly indicate that lack of equivalent controls for question complexity could underlie contradictory findings for response time performance, and it may be inferred that similar effects could occur in accuracy performance, as observed between these two studies (Bettman and Zins 1979). Among experiments where differential levels of *both* question type and complexity are operationalized, our findings suggest that contradictory outcomes are *likely* to occur.

6.1 Approaches to Data Presentation Research

One purpose of this paper is to support future development of grounded data presentation theory. The criteria for such theory are that it must both *explain* and *predict* the phenomena that researchers observe (Bacharach 1989) and that it be *grounded* through integrating preceding empirical research (Glaser and Strauss 1967). Currently, in the few instances where theory exists in the data presentation literature, it exists on a piecemeal basis — for example, cognitive fit theory's prediction and explanation of display format by question type interactions.

This lack of over-arching theory, coupled with a long history of equivocal findings, has led to substantial skepticism concerning the future prospects of data presentation research. Lohse, for example, advocates replacing traditional empirical research with "a quantitative cognitive

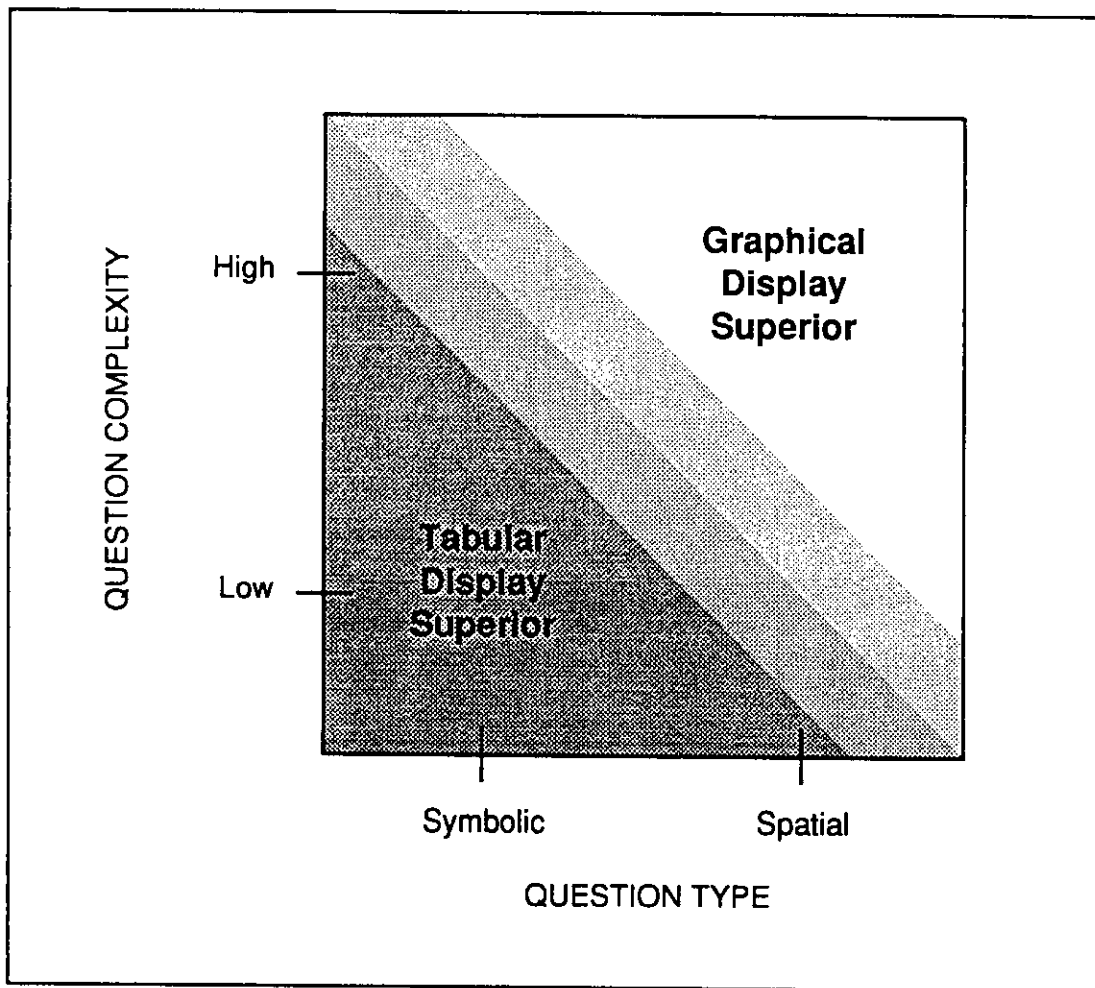


Figure 4. Pattern of Superior Display Format Performance Corresponding to Question Type and Question Complexity Task Characteristics

modeling approach [that] can provide robust objective predictions" (1993, p. 222). Such an approach has practical potential for predicting *what* display format to use within certain domain constraints, but it falls short of explaining general principles that govern *how* and *why* displays should be chosen. For this reason, we suggest that cumulative factor-based research still provides the best long-term approach to the goal of understanding how data presentations support the processes by which information is acquired from them. We subscribe to the view that it will be possible through a program of continued cumulative study to integrate the literature, including the findings of this study, into grounded data presentation theory that is both predictive and explanatory.

6.2 Limitations of the Research

This research is limited in several ways. First, it was not possible to include within a single study all the variables

that may influence performance through either main effects or interaction. Second, the investigated variables were operationalized at only two levels. In the case of question type, this was a theory-driven decision. It is clear, however, that display format and question complexity variables could be investigated with near-unlimited differentiation. In the case of display format, we chose a graph type (bar graph) which has been used in related research; however, this approach is limited in that it does not address the issue of comparing performance among graphical formats. Question complexity in this study was set at low and high levels, but these levels were so labeled in the interest of comprehensibility rather than as endpoints of the potential range of complexity. It is clear that much higher levels of question complexity — as well as an infinite number of intermediate levels — could be implemented, and it is uncertain how well the findings of this study will generalize to these levels.

6.3 Implications for Research and Practice

The findings have implications for researchers in their design and interpretation of data presentation studies. Many authors have criticized the existing data presentation literature, as a body, for using designs that are simplistic and lacking in controls (e.g., DeSanctis 1984; Jarvenpaa, Dickson, and DeSanctis 1985; Carswell 1992). Our findings specifically prescribe that increased attention be paid to interactions that exist between display and task variables. Factorial designs which test for two-way interactions are preferable to simple testing of group means. However, it is evident in this research domain that joint effects of interaction must be considered in order to capture important dimensions of the phenomena. This said, it does not follow that studies which initially failed to find encompassing "tables versus graphs" effects should be dismissed out of hand. On the contrary, equivocal research streams may provide valuable direction for further research, and sources for meta-analytic study, when effects of variable interactions are considered.

There are also implications for data presentation designers and users. Our findings show that tables provide good support not only for tasks that require precise answers but also for those that involve simple comparisons. When the supported tasks have these characteristics, the designer who uses graphs for every purpose gives up accuracy and, frequently, speed as well. Conversely, these findings do support the idea that graphs provide an abstracting function that speeds up the process of comparing among data components at higher levels of question complexity.

7. ACKNOWLEDGMENTS

We thank the four anonymous ICIS referees for the depth and detail of their suggestions. The present paper benefits by incorporating numerous points that were raised during the review process.

8. REFERENCES

- Addo, T. B. A. *Development of a Valid and Robust Metric for Measuring Question Complexity in Computer Graphics Experimentation*. Unpublished Doctoral Dissertation, Indiana University, Bloomington, 1989.
- Addo, T. B. A. "The Effects of Dimensionality in Computer Graphics." *Journal of Business Communication*, Volume 31, Number 4, 1994.
- Bacharach, S. B. "Organizational Theories: Some Criteria for Evaluation." *Academy of Management Review*, Volume 14, Number 4, 1989, pp. 496-515.
- Bertin, J. *Semiology of Graphics* (W. J. Berg, trans.). Madison, Wisconsin: University of Wisconsin Press 1983 (Original work published 1967).
- Bettman, J. R., and Zins, M. A. "Information Format and Choice Tasks in Decision Making." *Journal of Consumer Research*, Volume 6, Number 2, 1979, pp. 141-153.
- Carswell, C. M. "Reading Graphs: Interactions of Processing Requirements and Stimulus Structure." In B. Burns (Editor), *Percepts, Concepts and Categories*. North Holland: Elsevier Science Publishers, 1992, pp. 605-645.
- Coll, J. H. "An Experimental Study of the Efficacy of Tables versus Bar Graphs with Respect to Type of Task." *Information and Management*, Volume 23, 1992, pp. 45-51.
- Davis, L. R. *The Effects of Question Complexity and Form of Presentation on the Extraction of Question-Answers from an Information Presentation*. Unpublished Doctoral Dissertation, Indiana University, 1986.
- DeSanctis, G. "Computer Graphics as Decision Aids: Directions for Research." *Decision Sciences*, Volume 5, Number 3, 1984, pp. 463-487.
- Dickson, G. W.; Senn, J. A.; and Chervany, N. L. "Research in Management Information Systems: The Minnesota Experiments." *Management Science*, Volume 23, Number 9, 1977, pp. 913-923.
- Glaser, B. G., and Strauss, A. L. *The Discovery of Grounded Theory*. Chicago, Illinois: Aldine, 1967.
- Hackman, J. R. "Toward Understanding the Role of Tasks in Behavioral Research." *Acta Psychologica*, Volume 31, 1969, pp. 97-128.
- Hays, W. L. *Statistics for Psychologists*. New York: Holt, Rinehart, and Winston, 1963.
- Hoadley, E. D. "Investigating the Effects of Color." *Communications of the ACM*, Volume 33, Number 2, 1990, pp. 120-125.
- Ives, B. "Graphical User Interfaces for Business Information Systems." *MIS Quarterly*, Special Issue, December 1982, pp. 15-47.

- Jarvenpaa, S. L. "The Effect of Task Demands and Graphical Format on Information Processing Strategies." *Management Science*, Volume 35, Number 3, 1989, pp. 285-303.
- Jarvenpaa, S. L.; Dickson, G. W.; and DeSanctis, G. "Methodological Issues in Experimental IS Research: Experiences and Recommendations." *MIS Quarterly*, Volume 9, Number 2, 1985, pp. 141-156.
- Johnson, J. R.; Rice, R. R.; and Roemmich, R. A. "Pictures that Lie: The Abuse of Graphs in Annual Reports." *Management Accounting*, Volume 62, Number 4, 1980, pp. 50-56.
- Joyner, E. R. *The Development of a Metric for Assessing the Information Complexity of Time-Series Business Graphs*. Unpublished Doctoral Dissertation, Indiana University, Bloomington, 1989.
- Lauer, T. W. *The Effect of Variations in Information Complexity and Form of Presentation on Performance of an Information Extraction Task*. Unpublished Doctoral Dissertation, Indiana University, Bloomington, 1986.
- Leeuwenberg, E. L. J. *Structural Information of Visual Patterns: An Efficient Coding System in Perception*. The Hague: Mouton and Co., 1968.
- Lohse, G. L. "Eye Movement-Based Analysis of Graphs and Tables: The Next Generation." In J. I. DeGross, R. P. Bostrom, and D. Robey (Editors), *Proceedings of the Fourteenth International Conference on Information Systems*, Orlando, Florida, 1993, pp. 213-224.
- Mandeville, G. K. *A Monte Carlo Investigation of the Adequacy of Standard Analysis of Variance Test Procedures for Dependent Binary Variates*. Unpublished Doctoral Dissertation, University of Minnesota, Minneapolis, 1969.
- Mandeville, G. K. "A New Look at Treatment Differences." *American Educational Research Journal*, Volume 9, Number 2, 1972, pp. 311-321.
- O'Brien, R. G., and Kaiser, M. K. "MANOVA Method for Analyzing Repeated Measures Designs: An Extensive Primer." *Psychological Bulletin*, Volume 97, Number 2, 1985, pp. 316-333.
- Pachella, R. G. "The Interpretation of Reaction Time in Information-Processing Research." In B. H. Kantowitz (Editor), *Human Information Processing: Tutorials in Performance and Cognition*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1974, pp. 41-82.
- Parkman, J. M. "Temporal Aspects of Digit and Letter Inequality Judgments." *Journal of Experimental Psychology*, Volume 91, 1971, pp. 191-205.
- Remus, W. E. "A Study of Graphical and Tabular Displays and Their Interaction with Environmental Complexity." *Management Science*, Volume 33, Number 9, 1987, pp. 1200-1204.
- Schwartz, D. R., and Howell, W. C. "Optional Stopping Performance Under Graphic and Numeric CRT Formatting." *Human Factors*, Volume 27, 1985, pp. 433-444.
- Tan, J. K., and Benbasat, I. "Processing of Graphical Information: A Decomposition Taxonomy to Match Data Extraction Tasks and Graphical Representations." *Information Systems Research*, Volume 1, Number 4, 1990, pp. 416-439.
- Taylor, B. G., and Anderson, L. K. "Misleading Graphs: Guidelines for the Accountant." *Journal of Accountancy*, October 1986, pp. 126-135.
- Vessey, I. "Cognitive Fit: A Theory-Based Analysis of the Graphs versus Tables Literature." *Decision Sciences*, Volume 22, Spring 1991, pp. 219-241.
- Vessey, I., and Galletta, D. "Cognitive Fit: An Empirical Study of Information Acquisition." *Information Systems Research*, Volume 2, Number 1, 1991, pp. 63-84.
- Vickers, D. *Decision Processes in Visual Perception*. New York: Academic Press, 1979.
- Wood, R. E. "Task Complexity: Definition of the Construct." *Organizational Behavior and Human Decision Processes*, Volume 37, 1986, pp. 60-82.
- Wright, P. "The Harassed Decision Maker: Time Pressures, Distraction, and the Use of Evidence." *Journal of Applied Psychology*, Volume 59, Number 5, 1974, pp. 555-561.
- Yoo, K. *The Effects of Question Difficulty and Information Complexity on the Extraction of Data from an Information Presentation*. Unpublished Doctoral Dissertation, Indiana University, Bloomington, 1985.
- Zmud, R. W.; Blocher, E.; and Moffie, R. P. "The Impact of Color Graphic Report Formats on Decision Performance and Learning." In C. A. Ross and E. B. Swanson (Editors), *Proceedings of the Fourth International Conference on Information Systems*, Houston, Texas, 1983, pp. 179-193.