Management of Computing

Gordon B. Davis
Editor

Understanding The Effectiveness of Computer Graphics for Decision Support: A Cumulative Experimental Approach

GARY W. DICKSON, GERARDINE DeSANCTIS, and D. J. McBRIDE

ABSTRACT: A total of 840 junior and senior-level undergraduate business students particpated in three experiments that compared computer-generated graphical forms of data presentation to traditional tabular reports. The first experiment compared tables and bar charts for their effects on readability, interpretation accuracy, and decision making. No differences in interpretation accuracy or decision quality were observed for the two groups, although tabular reports were rated as "easier to read and understand" than graphical reports. The second experiment compared line plots to tables for their effects on interpretation accuracy and decision quality. Subjects with graphical reports outperformed those with tables. There were no meaningful differences in interpretation accuracy across treatment groups. The third experiment compared graphical and tabular reports for their ability to convey a "message" to the reader. Only in situations in which a vast amount of information was presented and relatively simple impressions were to be made, did subjects given graphs outperform those using tables.

This program of cumulative experiments indicates that generalized claims of superiority of graphic presentation are unsupported, at least for decision-related activities. In fact, the experiments suggest that the effectiveness of the data display format is largely a function of the characteristics of the task at hand, and that impressions gleaned from "one shot" studies of the effectiveness of the use of graphs may be nothing more than situationally dependent artifacts.

1. INTRODUCTION

As computer graphics make their entrance into corporate reporting and decision support systems, many questions arise regarding the impact of this technology on the quality of organizational decision making. Overall, research on the effectiveness of graphs as decision support tools is rather sparse. Moreover, studies have shown highly conflicting results when graphs are compared to more traditional tabular methods of data presentation (see [5, 11, 14]). These findings are not very encouraging in light of the many advantages of business graphics espoused by vendors and others in the popular media. Although some studies show graphs to be superior to tables [2, 3, 6, 19], others report the opposite conclusion [7, 10, 13, 15, 20, 21]. One potential explanation for these confusing results may be the variability of task environment across research settings. It may be that the ability of graphics to serve as effective decision aids depends on the specific decision context, or type of decision-related task.

A series of experiments was conducted to test one general hypothesis: that the effectiveness of computer graphics as a decision support tool varies as a function of the task environment in which the user is operating. Scanning the decision making literature, one is struck by the fact that "task environment" is not an especially well understood concept. What the literature does tell us is that "task" is very important in decision making [16] and that several dimensions of "task" appear to play a role in decision making. Among these factors are: task content, task complexity, and degree of task structure (e.g., see [9, 16]). There is no agreement on the definitions of these factors, or on what methods should be used to operationalize them. Given the conflicting evidence on the effectiveness of graphics in decision support cited above, a series of studies was conducted to explore the role of task environment in decision making. Task content, task complexity, and task structure were different for each study. Our objective was not to compare the importance of these three factors with one another, but rather to investigate the *overall* role of task environment in the use of graphics.

Three laboratory experiments were conducted in which a total of 840 junior and senior-level business students participated. All three experiments compared the decision supporting effectiveness of two media, tables and graphs, in terms of interpretation accuracy and decision quality. In the first experiment, low task structure and low task complexity were involved. In the subsequent two experiments, both task structure and task complexity were increased to test the importance of these constructs in determining the impact of graphics on decision performance. The reader should note the increasing sophistication in the design of the studies over time.

EXPERIMENT 1

Purpose and Background

This experiment compared tables and graphs for their effects on data interpretation accuracy and decision making effectiveness in a financial/accounting context. For this first experiment, we picked relatively simple levels of task content, task complexity, and task structure. The task content was an accounting setting that would be very familiar to all of the subjects. Task complexity was very low in that the subjects only had to process one variable at a time to reach a decision. The task was highly structured in that the subjects were given a step-by-step procedure to use when reaching a decision.

It has been argued that tasks of this sort are better supported by graphs than tables [8, 18]. Our choice of dependent variables for this initial experiment was influenced by the research of [1, 4, 19] who studied the effectiveness of graphs relative to tables.

Procedure

One-hundred fifty-four subjects participated in this experiment. The subjects were presented with a short case which described the situation of a small business in need of a loan. Subjects were told that they were to play the role of a bank loan officer. They were asked to read financial statements of the small business (balance sheet, income statement, return on assets, and return on investment reports), determine if the firm qualified for a loan, determine the maximum amount of the loan, and rate the riskiness of the loan. Half of the subjects received tabular reports; the remainder received graphical reports. Tabular information was displayed in the manner traditional to standard financial reports. Graphs were prepared according to recommendations made by [12]. Jarett's work is the only source of instruction currently available for preparing graphical representations of standard accounting information. The balance sheet and income statement were shown with horizontal bar charts; the ROA and ROI data were shown as double bars in a vertical bar graph. Exact dollar figures were given at the end of the horizontal bar graphs so that users did not have to interpolate data values. (This fact and others associated with how data were displayed should be noted by the reader throughout this article, especially in regard to interpreting results associated with "accuracy.") Following the set of financial reports, subjects were given a "Loan Information Form" which included interpretive questions about the reports and a structured procedure for determining loan qualification, loan amount, and loan riskiness. Interpretive questions included the following:

> "What was the company's net income for the past year?"
> "Did sales exceed the cost of goods sold?"
> "Which asset had greater value, fixed assets or inventory?"

The structured methods for determining loan qualification, loan amount, and riskiness were simple and straightforward. For example, the procedure to decide loan amount was as follows:

> "The bank will lend the company an amount equal to 80 percent of Accounts Receivable, plus 50 percent of the value of inventory. If there are notes Payable, these must be subtracted from the total loan amount. What is the maximum amount the bank can loan to this company?"

Following completion of the Loan Information Form, the subjects were asked to rate the difficulty of the task and the readability of the reports on 7-point Likert scales.

Results

All items in the Loan Information Form were scored as either correct or incorrect. Scores for items requiring the subject to identify values, compare values, or observe trends were summed to create an "interpretation accuracy" score. Items related to loan qualification, loan amount, and loan riskiness were summed to create a "decision quality" score. Mean scores for the tabular and graphic treatments were compared using *t*-tests. Results are shown in Table I. Interpretation accuracy and decision quality scores were not significantly different for the two groups. Subjects receiving graphical reports tended to rate the task as more difficult, but differences in ratings between the two treatment groups were not significant. With regard to report readability, the graphical group found the reports to be more difficult to read than the tabular group, and this difference was significant.

These results suggest that for simple financial reporting, use of graphical displays will neither hinder nor improve interpretation accuracy or decision quality over that observed with traditional tabular reports. Further, users appear to prefer a report format with which they are familiar over a novel method of data display.

EXPERIMENT 2

Purpose and Background

This experiment compared tabular and graphical methods for their effects on data interpretation and decision making for a market demand forecasting problem. In this experiment, the task content dealt with forecasting demand for a product. This context, while generally familiar to the subjects, was felt to be slightly more sophisticated than working with the basic financial statements required in Experiment 1. The problem was more "unstructured" than that of Experiment 1 in that no guidelines were provided to the subjects as to how to forecast the demand. Finally, whereas the first experiment involved processing a single variable at a time, Experiment 2 required consideration of two variables, level of demand and time (thus increasing the level of task complexity). Experiment 2 was anchored to Experiment 1 and the work of others (e.g., [13, 23]) in its use and measurement of interpretation accuracy and decision quality as dependent variables.

TABLE I. *T*-tests of Differences Between Means Experiment 1

Dependent variable	Tabular group	Graphic group	*t*-value	Probability
Interpretation accuracy	6.43	6.48	.49	.626
Decision quality	7.47	7.39	.47	.642
Task difficulty	2.41	2.52	.60	.546
Report readability	1.75	2.05	2.29*	.023

* Significant *t* value

Procedure

Three-hundred twenty students participated in this experiment. None of these subjects had participated in Experiment 1. Subjects were first presented with a short case describing a chemical manufacturer in need of assistance in forecasting demand for three of its products. The subjects were then given demand histories for each of the three products. Half of the subjects received graphical reports; the remainder received tabular reports. Each product report contained a 30-month demand history. Historical demand data for the products were generated using the following three sine functions:

Product	*Demand pattern*
"Resinoid"	Negatively sloping linear function with little variability
"Resinforced"	Cyclical function with some variability
"Vitrified"	Cyclical function with much variability

Notice that the functions varied in complexity; the first was the simplest and the third was the most difficult to forecast. After reading the three reports, the subjects were asked five interpretive questions. Examples are as follows:

"What was the demand (in pounds) of the Resinoid product in month 20?"
"What was the difference in demand between month 1 and month 10 for the Vitrified product?"
"Compare the demand pattern for the Resinoid product with the demand pattern for the Vitrified product. Explain how these patterns differ from one another."

Following the interpretation questions, subjects were asked to provide specific estimates of demand for each of the three products for three months into the future. Finally, they were asked to rate the difficulty of the task on a 5-point Likert scale.

Results

The items in the experimental questionnaire were scored by determining the absolute value of the difference between the subject's response and the correct answer to the question. Scores for the five items dealing with data interpretation were summed to create an "interpretation accuracy" score. The average interpretation accuracy scores for both the tabular and graphic groups were reasonably low (lower scores indicate better performance). Scores for the graphic group were lower (mean = 2414) than those for the tabular group (mean = 4000)—a surprising finding in view of the fact that the exact data values should have been more difficult to identify on the graphs than on the tables (fairly

TABLE II. *T*-Tests Comparing Tabular and Graphic Treatments for Nine Dependent Variables Experiment 2

Dependent measure	Tabular	Graphic	*t*-value	Probability
Resinoid 1	2333	1720	3.34*	.001
Resinoid 2	2092	1395	2.69*	.007
Resinoid 3	2785	1666	4.45*	.000
Resinforced 1	3776	2264	3.92*	.000
Resinforced 2	4162	2527	2.65*	.009
Resinforced 3	4269	2971	3.23*	.001
Vitrified 1	4134	2691	2.83*	.005
Vitrified 2	3872	2310	3.64*	.000
Vitrified 3	3348	2902	1.49	.137

* Significant *t* value

Note: All measures are taken as the absolute value of the difference between the subject's forecast and the correct forecast. Therefore, larger numbers indicate greater distance from the correct forecast.

imprecise triangles were used to represent data points on the graph). However, the difference between interpretation scores for the two groups was not significant ($t = 2.21$, $p = .07$).

"Decision quality" was measured in terms of the forecast accuracy in each of three periods for each of three products. *T*-tests comparing the performance of the two groups are shown in Table II. In eight of the nine forecasts, subjects presented with graphs outperformed those working with tables. It is most appropriate to examine performance using the first period demand estimate for each of the hypothetical products, that is, "Resinoid 1," "Resinforced 1," and "Vitrified 1." These are the best measures of "decision quality" in that later period forecasts tend to exaggerate any error made by the subject in the first period. As shown in Table II, the graphic users outperformed the subjects given tabular data in all three of these forecasts. Consistent with their better performance, the graphical group perceived the task to be easier than the tabular group ($F = 22.2$, $p = .001$).

In contrast with the results generated in Experiment 1, data presentation using graphs enhanced "decision quality" in Experiment 2. Considered in combination, these results imply that it may not be appropriate to generalize about the superiority of a particular method of data presentation across task environments.

EXPERIMENT 3

Purpose and Background

This experiment compared tables and graphs for their effects on the ability of decision makers to "get the message" from material in a general managerial context. Having performed Experiments 1 and 2, we again increased the overall complexity of the task environment under investigation. First, we shifted the task content to involve the subject area of business graphics. Whereas in the previous two experiments the subjects had some familiarity with the task area, few would be very familiar with the "state of the art" in business graphics. Second, the task was highly unstructured in that the subjects were asked a number of questions based on the material presented to them but no guidelines were provided as to how to perform this task; in contrast to Experiments 1 and 2, there was no historical basis for forming a judgment. Third, this task was more complex than the prior two in that it required the subject to simultaneously consider a number of variables rather than only one or two at a time.

As in the previous two experiments, information accuracy and quality of decision making were important dependent variables. The works of [17, 20, 22] were used to develop measures of these variables for this study. The experimental situation was more complex than either described in Experiment 1 or Experiment 2 in that the tabular versus graphic treatments were examined under four combinations of conditions.

Procedure

Three-hundred sixty-three students participated in this experiment. None of these subjects participated in Experiments 1 or 2. In total, nine treatment groups were employed. Subjects were first presented with a short case describing a producer of computer graphics software that had contracted with a research organization to do a survey of users of computer graphics. The subjects were told that they were involved in an experiment to evaluate the quality of the research firm's final report. They were told that they would receive a report on current usage of graphics technology in business and would be asked a series of questions about what the report was trying to convey.

A 2 × 2 × 2 factorial design was used. The independent variables were: (1) tabular presentation versus

TABLE III. Summary of Reports and Measures Experiment 3

Treatment	Number of reports	"Message" measure questions	"Traditional" measure questions	Total questions
Complete	9	32	17	49
Subset				
Number 1	4	18	6	24
Number 2	5	14	11	25

graphic presentation, (2) complete presentation versus subset presentation, and (3) recall versus lookup. Subjects in the "complete" treatment groups were given 9 graphs or tables and had to answer a total of 47 questions. Subjects in the "subset" groups did the task in two parts. They were first given 4 reports and answered 22 questions, and then answered 25 questions based on 5 reports. Subjects in the "recall" groups were told to look at the reports (graphs or tables) for a period of time (which was a function of whether they were in a complete or a subset group), put the set of reports away, and answer the questions. Subjects in the "lookup" groups had the set of reports available to them as they were responding to the questions.

Two different measures were employed to determine the extent to which subjects received the main messages contained in the reports. The first measure was based on answers noting agreement or disagreement with statements describing the "messages" given in the reports. Two examples of statements used were:

"Larger organizations tend to be the major users of business graphics."
"No mainframe graphics package dominates over others."

Of the 30 statements of this type, some were true and were major messages from the reports. Others were incorrect statements of two types; either they dealt with material in the reports but were wrong, or they were statements that might be true or false but were not covered in the reports given to the subjects.

The second measure was more "traditional" in nature and resembled interpretation accuracy measures used in other graphics studies. An example of a question used to construct this measure was:

"How does a 'vendor offer' rank among the factors influencing the purchase of graphics technology?

a. First c. Third
b. Second d. Last"

All questions used in constructing this measure were related to material in the reports given to the subjects. Table III summarizes the number of reports and types of questions given to the various treatment groups. Questions were ordered randomly within both the "complete" and "subset" treatment groups to control for order effects. As a result of the randomization, it was possible that subjects in the "lookup" treatment groups paged back and forth looking for the material to answer a question.

Fixed amounts of time were allotted for examining reports and answering questions. These time periods varied, of course, across "lookup" and "recall" conditions as well as "complete" and "subset" conditions. However, every attempt was made to make the timeslots comparable across treatment groups. In every case, 45 minutes were available for the total experiment.

One additional group was employed in this study. This group served as a baseline to ensure that the performance on the task was a function of reading the reports and not of general knowledge. The baseline group answered all 47 questions without seeing any reports at all. On the average, the baseline group scored 7.03 points less than the poorest performing treatment group on the "message" measure (out of 30 possible points) and 4.98 points less than the poorest performing treatment group on the "traditional" measure (out of 17 possible points); this confirmed that performance on the task required the subjects to read the reports.

Results

The measure for "getting the message" was scored by summing the number of correct and incorrect statements appropriately identified in a list of 32 statements. The more "traditional" message measure was simply the total number of correct answers to a set of 17 questions. Table IV shows the results for the "message" measure, and Table V depicts the same information for the "traditional" measure.

A few salient observations can be made simply from scanning these tables. First, on the traditional measure, the lookup groups did very well (scoring almost perfectly), while the recall groups did poorly. On the other hand, there was little difference in performance on the message measure. About the only clear pattern to emerge is that the groups getting the information in "chunks" (the subset groups) nearly always "got the message" better than their counterparts working with the entire set of reports.

In support of these observations, Tables VI and VII are offered. We see from Table VI that, for the message measure, there is a significant subset effect. None of the two-way interaction effects are significant. Table VII shows a very different picture, however. In this case, there are two significant two-way interactions, format by subset and visual by subset.

These results suggest that it made a difference to the recall group (but not the lookup group) whether or not they had the complete set or two subsets of reports. The recall group performed better with subsets. Similarly, performance on the traditional measure showed a relationship between the format employed and whether or

TABLE IV. Comparison of Mean Scores on the "Message" Measure—Experiment 3

	Tabular		Graphic	
	Lookup	Recall	Lookup	Recall
Complete	22.22	23.21	23.38	23.00
Subset	24.38	24.10	24.18	23.29

TABLE V. Comparison of Mean Scores on the "Traditional" Measure—Experiment 3

	Tabular		Graphic	
	Lookup	Recall	Lookup	Recall
Complete	16.10	9.77	16.24	10.36
Subset	16.42	11.58	15.97	11.22

not the subjects had a complete set of reports or a subset. In the case of having the complete set of reports (and many questions to answer), the graphical group did significantly better than the tabular group. However, in the case of those receiving the material in two parts, the opposite result occurred.

From this experiment, the researcher can get several "messages." First, if presenting a large amount of material to managers, to make sure that major messages are conveyed, it would be better to break up the material than to do almost anything else. Second, there is evidence that what is measured concerning subject performance (and probably how it is measured), can affect the results obtained and the story passed on to both the research and practitioner communities. A third message from this experiment is that the use of graphs showed an advantage over tables for enhancing decision performance in one instance. This was the situation in which a large amount of material was presented (complete group) with relatively few impressions to be gleaned (as represented by the traditional measure).

As a final observation, we would like to note that while time was a controlled variable in the study, time *pressure* was not. The total amount of time provided to read and respond to reports was the same across groups. However, when conducting the experiment it appeared that the time pressure experienced by the subjects varied across groups. It is possible that the results for the two message measures would have been more similar had the amounts of time allotted to the two measures been balanced, rather than equal, so that equivalant time pressure would be experienced by the subjects. Determining what constitutes time equivalency in an experiment like this would be a difficult task.

DISCUSSION

The set of experiments described in this article presents a first-level look at the role of task environment in the effectiveness of graphs as decision support tools. In addition to supporting the original contention of [16] and the present authors that the effectiveness of decision aids is influenced by the task in which they are employed, we feel that a good deal of insight has been gained into both the use of graphs and conducting additional research in this area. A relatively simple experi-

TABLE VI. ANOVA Results for the "Message" Measure Experiment 3

	F	Probability
Main Effects		
Format (tables versus graphs)	.001	.971
Visual (lookup versus recall)	.145	.704
Subset (complete versus subset)	8.576*	.004
Interaction Effects		
2 Way		
Format × Visual	2.040	.154
Format × Subset	1.927	.166
Visual × Subset	1.613	.205
3 Way		
Format × Visual × Subset	2.978	.586

* Significant *F* value

TABLE VII. ANOVA Results for the "Traditional" Measure Experiment 3

	F	Probability
Main Effects		
Format (tables versus graphs)	.001	.979
Visual (lookup versus recall)	676.934*	.001
Subset (complete versus subset)	11.048*	.001
Interaction Effects		
2 Way		
Format × Visual	.415	.520
Format × Subset	3.408*	.066
Visual × Subset	9.688*	.002
3 Way		
Format × Visual × Subset	.100	.665

* Significant *F* value

ment was conducted first, and from the results of that experience a more complex but complementary study was designed and performed. Following these two experiments, an even more sophisticated study was conducted. By taking this approach, the researchers can now continue to expand the work further or to focus in on a few detailed questions of great interest and potential promise. We intend to do both.

Several interesting patterns emerge from the results of these studies. Perhaps the most important implications are for further research. We will return to this topic shortly. First we will present the more practical messages gained from our work. The authors want to emphasize that the discussion which follows only relates to the use of graphics in decision related tasks and that other uses of graphs, such as to influence action or persuade an audience to accept a point of view may be an entirely different situation.

IMPLICATIONS FOR THE USERS OF BUSINESS GRAPHICS

Our studies do not offer much "good news" for the vendors and users of business graphics. The best news for these parties is that using graphs is no worse than using tables. We have identified a few specific situations in which graphs appear to be more effective than tables. However, claims of total superiority of graphs should be viewed with skepticism. In particular, we offer the following advice:

1. In tasks for which accurate interpretation of values is important, especially in environments in which the user has experience with tabular presentations (such as dealing with financial data), tables are probably a better choice of format than graphs.
2. For a task activity that involves seeing time dependent patterns in a large amount of data, graphs are a good choice of format.
3. If the purpose is to convey a message to an audience from several sets of data on a common subject, the best thing to do is "chunk" the data by breaking it into smaller sets. Either graphs or tables may be used.
4. Graphs may be the format of choice in a situation where people are presented with a large amount of data and the goal is to have them recall some fairly specific facts about the data immediately after the presentation.

We must exercise caution in accepting these guidelines. Clearly, we need to recreate the circumstances we have defined and attempt to replicate the results that underlie our tentative conclusions.

The preparation of graphs in support of decision analysis can be expensive and time consuming. Other than instances in which the goal is to spot trends in the data, these researchers see little to recommend going to the trouble of creating graphs for this application. We realize that this is a strong statement and must be tempered with several disclaimers. One is that our conclusions are restricted to decision applications of graphics. Another is that we have not used subjects experienced and familiar with the use of business graphs nor have we provided training in the use of these tools. A third is that we must replicate our findings regarding task activities by subjecting graphs to even more tests in task environments with properties similar to those upon which our tentative conclusions are based. A final disclaimer is that our observations are based on a set of single-period experiments, and it is not unreasonable to expect that a repeated measures design (to accommodate learning) could yield different results.

IMPLICATIONS FOR RESEARCHERS

For persons wishing to do research on business graphics, our work allows us to make several suggestions.

1. Doing experimental research on the effectiveness of business graphics virtually demands that a cumulative and programmatic approach be taken. One must iteratively investigate a situation, continually sharpening measurement procedures and confirming results in a particular task environment before moving on to totally new settings. On the latter point, researchers might note that we chose only three areas (interpretation accuracy, trend spotting, and transmission of major messages) to explore rather thoroughly. Our intent is to continue to work on measurement and replication in these task settings, as well as to move ahead in promising new directions.
2. Note should be taken of how influenced our results are by the task activity and, as shown in Experiment 3, of how the dependent variable was measured. Researchers must work hard on developing and testing reliable and valid outcome measures. It is clear from our work that the message one takes away from a study is very dependent upon how outcomes are measured.
3. For this kind of exploratory, "theory building" research, the use of student subjects is very appropriate. In the first place, for activities of the type our experiments require, there is no reason to believe that students would perform any differently than managers. Additionally, for this kind of work, one needs the power in tests that can be obtained by the large numbers available in student subjects.
4. Additional research might explore in more depth the ability of graphs to "summarize" large amounts of data in order that users can get the messages contained in the data. Our analysis using the "traditional" measure showed one instance of this ability. Further exploration, perhaps with refined performance measures, appears warranted.
5. One area which looks very promising for further research is the use of graphics to persuade an individual or group to take action or accept a point of view. In our opinion, the effectiveness of graphics in this area may be much better than for decision support.

The series of experiments described in this article were undertaken with the purpose of finding out whether or not the effectiveness of computer graphics as a decision support tool varies as a function of the decision task environment in which the user is operating. Our evidence is very much in support of the proposition that the effectiveness of business graphics is a function of the task environment. This series of studies did not attempt to separate out the effects of "task content," "task complexity," and "task structure." The results hint that graphs outperform tables when the latter two factors are set at a moderate level (as opposed to very low or very high levels). More refined examination of the dimensions of task environment and their relationship to the use of graphics is a subject for future research.

CONCLUSIONS AND FUTURE DIRECTIONS

From the results we have obtained thus far in our program of research, it is easy to understand the conflicting results that have been generated by previous studies of the effectiveness of business graphics. Enhancements to our understanding of business graphics as an information systems tool can come about only by taking a directed and programmed research approach, in contrast to a piece-meal approach. The desired approach is obviously a time consuming process.

We intend to follow the guidelines presented in this article in our future research endeavors. We will revisit the literature to see if the tentative conclusions we have reached can be supported when prior work is viewed from the task activity framework we have begun to construct. We will seek out those studies that have task activities with properties similar to those we have investigated and look for confirming patterns in results. We will continue to explore the problem of measurement in our existing task settings and perform new experiments in settings with similar task activities in an attempt to confirm our beliefs. Our research will also attempt to separate out the effects of what we have called "task content," "task complexity," and "task structure." We will also move to the use of graphs in the context of persuasion. Again, we will adopt a programmatic approach in our research, starting with fairly simple studies and then moving on to complex designs as our understanding is confirmed. Finally, our research team will explore the effect of learning by undertaking studies involving repeated measure designs. Over a time horizon of two years, we expect to conduct 8–10 experiments within a program devoted to the study of managerial graphics.

REFERENCES

1. Benbasat, I., and Dexter, A.S. An experimental evaluation of graphical and color-enhanced information presentation. *Manage. Sci.*, to be published.
2. Benbasat, I., and Schroeder, R. An experimental investigation of some MIS design variables. *MIS Q. 1*, 1 (1977), 37–50.
3. Carter, L.F. An experiment on the design of tables and graphs used for presenting numerical data. *J. Appl. Psychol. 31*, (1947), 640–650.
4. Carter, L.F. Relative effectiveness of presenting numerical data by the use of tables and graphs. U.S. Department of Commerce, Washington, D.C., 1948.
5. DeSanctis, G. Computer graphics as decision aids: Directions for research. *Decis. Sci. 15*, 4 (1984), 463–487.
6. Feliciano, G.D., Powers, R.D., and Bryant, E.K. The presentation of statistical information. *Audio Visual Commun. Rev. 11*, 13 (1963), 32–39.
7. Ghani, J.A. The effects of information representation and modification of decision performance. Ph.D. dissertation, Univ. of Pennsylvania, Philadelphia, 1981.
8. Ghani, J.A., and Lusk, E.J. Human information processing research: Its MIS design consequences. *Hum. Syst. Manage. 3*, 1 (1982), 32–40.
9. Gorry, G.A., and Scott-Morton, M.S. A framework for management information systems. *Sloan Manage. Rev. 13*, 1 (1971), 55–70.
10. Grace, G.L. Application of empirical methods to computer-based system design. *J. Appl. Psychol. 50*, (1966), 442–450.
11. Ives, B. Graphical user interfaces for business information systems. *MIS Q.*, Special Issue, (1982), 15–47.
12. Jarett, I.M. *Computer Graphics and Reporting Financial Data.* Wiley, New York, 1983.
13. Lucas, H.C. An experimental investigation of the use of computer-based graphics in decision making. *Manage. Sci. 27*, 7 (1981), 757–768.
14. MacDonald-Ross, M. How numbers are shown. *Audio-visual Commun. Rev. 25*, 4 (1977), 359–409.
15. Nawrocki, L.H. Alphanumeric versus graphic displays in a problem-solving task. *Tech. Res. Note 227*, U.S. Army Behavior and Systems Research Laboratory, Arlington, Va., Sept. 1972.
16. Payne, J.W. Contingent decision behavior. *Psychol. Bull. 92*, (1982), 382–402.
17. Powers, M., Lashley, C., Sanchez, P., and Schneiderman, B. An experimental comparison on tabular and graphic data presentation. Computer Science Department, University of Maryland, College Park, MD, Feb. 1982, unpublished.
18. Takeuchi, H., and Schmidt, A.H. New promise of computer graphics. *Harvard Bus. Rev. 58*, 1 (1980), 122–131.
19. Tullis, T.S. An evaluation of alphanumeric, graphic, and color displays. *Hum. Factors* 23, 5 (1981), 541–550.
20. Vernon, M.D. The use and value of graphical material in presenting quantitative data. *Occup. Psychol. 26*, (1952), 22–34.
21. Wainer, H., and Reiser, M. Assessing the efficacy of visual displays. In *Proceedings of the American Statistical Association, 1*, Social Statistics Section, (1976), pp. 89–92.
22. Wilcox, W. Numbers and the news: Graph, table or text. *Journalism Q. 41*, (1964), 38–44.
23. Zmud, R.W., Blocher, E., and Moffie, R.P. The impact of color graphic report formats on decision performance and learning. In *Proceedings of the Fourth International Conference on Information Systems*, Society for Information Management, Chicago, Ill., 1983.

CR Categories and Subject Descriptors: H.1.2 [**Models and Principles**]: User/Machine Systems—*human factors*; H.4.2 [**Information Systems Applications**]: Types of Systems—*decision support*; I.3.1 [**Computer Graphics**]: General

General Terms: Experimentation, Management

Additional Key Words and Phrases: decision support, graphics, task setting

Received 8/85; revised 10/85; accepted 10/85

Authors' Present Addresses: Gary W. Dickson, Gerardine DeSanctis, and D.J. McBride, Department of Management Sciences, School of Management, 271 19th Avenue South, University of Minnesota, Minneapolis, MN 55455.