ICIS 1985 Proceedings | International Conference on Information Systems (ICIS)

1985

An Investigation of the "Tables Versus Graphs" Controversy in a Learning Environment

Gerardine DeSanctis
University of Minnesota

Sirkka L. Jarvenpaa
University of Minnesota

An Investigation of the "Tables Versus Graphs" Controversy in a Learning Environment

Gerardine DeSanctis
and
Sirkka L. Jarvenpaa

University of Minnesota
Management Sciences Department
271 19th Avenue South
Minneapolis, MN 55455
(612) 373-5211

ABSTRACT

The study of computer graphics as decision aids has become popular among MIS researchers in the last several years. However, this area of research, like many others in management information systems, has been plagued with methodological problems and contradictory findings. In light of these difficulties, the current study examined the "tables versus graphs" controversy within a learning environment. Seventy-five MBA students were exposed to one of three experimental treatments and asked to develop financial forecasts for fictitious companies over five experimental trials. Following their forecasts for each firm, participants were provided with feedback on the quality of their decisions. The information presentation treatments were as follows: (1) traditional spreadsheet (tabular), (2) graphs using "standard" scaling, and (3) graphs using "nonstandard" scaling. Results suggest that, although graphics may initially demonstrate no advantage over tables, they do show an advantage if decision makers are repeatedly exposed to the novel format and given feedback on their performance. Learning will occur even when improper scaling is used. The implication is that the effectiveness of graphics as decision aids depends on practice. Researchers are encouraged to employ repeated measures, or longitudinal, designs when examining the tables-versus-graphs controversy.

Introduction

How to best display data to decision makers has been a concern to MIS researchers since Mason and Mitroff (1973) first noted the importance of "presentation mode" in the design of information systems. A large portion of this research effort has centered on comparing the relative effectiveness of tables and graphs for the support of problem solving activities in business settings. Interest in "tables versus graphs" comparisons has intensified during the past few years as sophisticated, easy-to-use graphics technology has become incorporated into decision support systems. The underlying assumption in these studies is that graphics should facilitate clearer perception of data relationships and trends over tables.

The empirical research dealing with the effectiveness of graphs as decision aids has been quite controversial. Several studies have found graphs to be easier to interpret than tables; others have found the reverse; and still others report no difference in interpretation accuracy for the two display methods. Similar conflicting results have been found when graphs and tables are compared for their effects on interpretation speed, user preference, and decision confidence (see Ives, 1982; MacDonald-Ross, 1977). Of a total of 7 studies dealing with the impact of graphics on decision quality, only one reports graphs to be superior to tables; 3 conclude that tables are superior to graphs, and 3 have found no difference between the two formats (see DeSanctis, 1984).

Why is it that computer graphics are not proving to be more useful as tools for supporting decision making? Several investigators who have found graphs to be fairly ineffective in improving decision quality have postulated that learning must occur before graphical output becomes meaningful to people (e.g., Lusk & Kersnick, 1979; Vernon, 1946). Business data traditionally has been displayed in tabular form. Consequently, decision makers simply lack the experience needed to properly interpret novel formats. This argument implies that practice in viewing graphs might improve their meaningfulness to users and, over time, a performance advantage of graphs in contrast to tables could be observed. No empirical test of this hypothesis has been conducted, however. From an experimental design perspective, the tables-versus-graphs comparison needs to be made within a learning environment. That is, people should be given the opportunity to *practice* using graphical materials and be provided with *feedback* on their performance. To date, studies of presentation format have uniformly examined dependent measures at a single point in time rather than repeatedly over an extended time period.

Beyond the learning issue, an additional factor which has been discussed but not empirically examined for its importance in effective use of graphs is the degree of standardization across decision aids. The use of "standards" refers to the application of a set of predefined rules that direct the construction of a graph with regard to its components—such as size, color, shading, and scaling. Establishment of standards for graphs has been advocated as important for avoiding perceptual problems and subsequent misinterpretation of graphically portrayed data (Cox, 1978). Like statistics, graphs have the potential to "lie" about the data which they represent. For example, a poor choice of scaling can lead a reader to overlook significant variations in data values or cause "mountians to be made out of molehills." Guidelines for scaling and other components of graphs have been developed over the years, primarily by graphics artists and statisticians. However, the importance and validity of these standards have not yet been investigated—beyond "clinical" or casual observation. Moreover, it is unknown whether users can visually adapt to, for example, poor scaling if they encounter the problem on a regular basis. From a research design standpoint, studying standardization requires identification of proposed standards, followed by measurement of the relative impact on decision performance of graphs which conform to or violate those standards. The current study will focus on the role of standard *scaling* in effective use of graphics, because the existing graphics software provides great opportunities for users to violate scaling guidelines.

The purpose of this study is to determine the importance of scaling standards and learning in effective use of graphics as decision aids. Graphs which conform to and violate recommended scaling standards will be compared to each other and to traditional tables for their effects on decision quality. In a laboratory environment, a series of experimental trials will be conducted in order to determine if the effects of tables, standard graphs, and nonstandard graphs on decision making change as users are repeatedly exposed to these formats and given feedback on their performance. The following hypothesis will be examined:

H1 Effective use of graphics in decision making requires learning. (*a*: Graphs will initially show no advantage over tables for decision quality; however, the decision quality associated with graphs will be better than that with tables following practice. (*b*): Graphs with standard scaling will initially show an advantage over graphs with nonstandard scaling; however, the decision quality associated with nonstandard graphs will be as good as that with standard graphs following practice.

The study aims to avoid some of the methodological weaknesses noted in other graphical studies (Jarvenpaa et al., 1985) by (a) utilizing a valid, reproducible task, (b) assuring high quality graphical materials, and (c) using a multivariate approach to data analysis. In an effort to develop cumulative research, the study will build upon behavioral accounting research on the display of financial information.

In the next section we summarize literature in the areas of learning, graphical standards, and accounting and discuss their relationship to our hypothesis and experimental design. We next describe the research methodology, results, and implications for further study.

Supporting Literature

LEARNING

Learning is defined by psychologists as a relatively permanent change in behavior which occurs as a result of practice (Kimble, 1969). There are two aspects of human learning. Development of "declarative knowledge" means being able to recall or recognize information. A few studies have examined the impact of graphs on recall of information (Nawrocki, 1972; Watson & Driver, 1983). In addition, several experiments have asked subjects to recognize data points in forced-choice questions following exposure to graphs (Powers, et al., 1981; DeSanctis and Dickson, 1985). In all studies to date, recall and recognition instruments have relied exclusively on subjects' verbal responses when capturing declarative knowledge. Yet graphs also contain spatial knowledge. In order to asssess acquisition of both spatial and verbal declarative knowledge, the current study will ask participants to draw graphs after exposure, in addition to answering multiple-choice questions.

There has been much less attention given to the second aspect of learning, "procedural knowledge," in graphics research. Procedural knowledge consists of the processes that use, or apply, declarative knowledge. Whereas declarative knowledge is "static," procedural knowledge is a cognitive skill or the ability to perform intellectual procedures (Anderson, 1980). In the context of decision graphics, procedural knowledge refers to the skills of extracting patterns and relationships presented in graphs and applying the extracted information to an appropriate decision model or rule in order to reach an accurate decision. With the exception of the work of Olson (1975) on the acquisition of map-reading skills, there has been no empirical study of the procedural knowledge aspect of using graphics. Bettman and Zins (1979) have postulated that people may adapt format to task if given the time—even if the format is "poor" (such as with nonstandard scaling). Testing this kind of proposition is difficult because psychologists have done little research in procedural knowledge acquisition. The current study will only be a first step in this direction, as we examine behaviorally whether the procedural knowledge needed to use graphs in decision making can be acquired with practice.

STANDARDS

The need for the establishment of standards for graphics, including scaling, has been discussed by many authors. As early as 1916, Brinton demonstrated how easily people underestimate the relative importance of data if their judgments are based on "areas" rather than "points." Graham (1937) found that people tend to overestimate the length of vertical bar charts, and Vernon (1946) suggested that people tend to focus on the "raw picture" of a chart and ignore information in the titles and axes. If standards were applied to graphical scaling, this should alleviate perceptual problems, lower the time required for generating and reading graphs, and perhaps increase managerial acceptance of graphs as a formal reporting method in organizations. The current study will consider two scaling standards: round numbers as interval values on scales, and identical scales of measurement on related charts. A "nonstandard graphics" experimental condition will be created by violating these guidelines. The violation should cause the users of nonstandard graphs to overstate variances in data values and encounter difficulties in comparing data across charts.

ACCOUNTING

Accountants have long been interested in identifying effective methods of communicating financial information to potential investors, stockholders, and auditors. And recently there has been an interest in establishing standard methods of graphically presenting financial data. In studies of financial statement formats, traditional statements have been compared with novel formats—such as forecast formats (Brandon & Jarrett, 1977), extremely complex formats (Pratt, 1982), and aggregation formats (Otley & Dias, 1982). A popular experimental task has been to ask students (or others with minimal knowledge of accounting) to recall and/or forecast a company's earnings based on historical data contained in spreadsheets or comparative income statements. In some cases novel formats have proved surprisingly superior to traditional formats (e.g., Moriarity, 1979). But in other cases subjects have performed best, or equally as well, with traditional formats, even in cases where the novel formats should be superior theoretically (e.g., Brandon & Jarrett, 1977). Only one study has considered graphics as an alternative to traditional tables; Moriarity (1979) discovered that people can predict corporate bankruptcy better with Chernoff faces than with tables. According to Brandon and Jarrett (1977), "students are unfamiliar with formats that deviate from the traditional statement forms and tend to disregard much of the informational content [contained in nontraditional statement forms]" (p. 701). The implication is that learning may be important for effective use of nontraditional displays of standard financial information.

Methodology

SUBJECTS AND SAMPLING PROCEDURES

Seventy-five MBA students enrolled in an introductory MIS course participated in the study. As part of their course, the students were given a choice of participating in this project or an equivalent non-research project. All participants were in the second year of their graduate program, and approximately ⅓ of the sample were employed full-time. On average, the students were 29 years old and had 4.4 years of full-time work experience in a business or administrative position.

Subjects were randomly assigned to one of three treatment groups: (1) tabular, (2) standard graphics, and (3) nonstandard graphics. The experiment required two hours of the subject's time. Data was collected in small groups of 4-to-7 students each, over a four-week period.

THE EXPERIMENTAL TASK

The experimental task and the measure of "decision quality" were developed based on the work of Brandon and Jarrett (1977, 1979), Pratt (1982), and others (Benjamin & Strawser, 1974) who have studied display methods for financial statements. Several pilot studies were conducted in order to refine the task, procedures, and measurement instruments. A 2-page case writeup and a set of reports were constructed for the experimental task. The case defined all the financial terms that subjects encountered in the task. The task required the subjects to read historical income and earnings per share statements for each of 5 firms and develop forecasts of EPS for the five firms. A data set of historical sales volume and expense information for five "companies" was generated using conventional "monte carlo" techniques. Both sales volume and expense variables were based on a linear time-series containing an error term. The error term was calculated by multiplying a normal random number for each year times a fixed standard deviation. Sales volume data were used to determine revenue and cost of goods sold information for 21 years, or periods; expense data were generated for 21 years as well, and then net income and earnings per share figures were derived for each of the 21 periods. The initial 16 years were used as historical data, while the final 5 years comprised future data. 21 years of earnings data were generated in this way for five "companies." To assure equivalency across the five firms, the normal random error terms and relationships among variables in the simulation model were held constant across all companies. Only the initial values for the first period sales and expenses, the number of stockholder shares, and the price and cost charged per unit, were modified for each of the firms.

EXPERIMENTAL MANIPULATION

The experimental manipulations were the format used to display the earnings data and practice in viewing a particular format. Traditional spreadsheets were used in the "tabular" condition. Horizontal bar charts, with dollar values appearing at the end of each bar, were used in the two graphical treatments. Horizontal bars were chosen over other formats because graphics experts argue for their use for depicting financial trends and relationships (Jarett, 1983). Six graphs were required for each of the 5 companies. All graphs were prepared according to guidelines proposed by Jarett (1983) for displaying financial information. In the "standard graphics" condition, values on the x-axis were in round numbers, and each graph was scaled according to the maximum revenue value for that *company*. In the "nonstandard graphics" condition, values on the x-axis were in nonround numbers, and each graph was scaled according to the maximum dollar value for that *graph*. (In general, this led to longer bars on the nonstandard graphs than on the standard graphs, although both had the same dollar values labelled at the end of each bar.) The practice variable was operationalized by exposing subjects to five experimental trials. The set of reports for each company constituted an experimental "trial." The order of presentation of the five companies was randomized across subjects to control for "order effects."

MEASUREMENT OF DECISION QUALITY

The experimental task required the subject to examine 16 periods of historical data and then estimate revenue, cost of sales, other expenses, net income, and earnings per share for three years into the future for each of the five companies. For each company, the subject's forecasts of earnings per share were compared to those values generated by the monte carlo model (Accurate EPS) in the following manner:

$$\text{Forecast Error} = \sum_{n=1}^{3} \frac{|\text{Forecast EPS} - \text{Accurate EPS}|}{\text{Accurate EPS}}$$

$$\text{Average Forecast Error} = (\text{Forecast Error}) / 3$$

$$\text{Average Percent Forecast Error} = \text{Average Forecast Error} * 100$$

The average percent forecast error was used as a measure of the subject's decision quality. This measure of forecast error was selected to keep the research consistent with the work on which it builds.

PROCEDURE

Prior to the experimental task, subjects completed a research participation consent form, an agreement to keep the nature of the study confidential, and a personal background questionnaire. As a performance incentive, subjects were informed that prize money of $50, $35, $25, and $10 would be awarded to the top four decision makers on the experimental task. Following the case that described the experimental task, subjects were provided with the reports for each company. Twelve minutes were allotted to evaluate the reports for each firm and record forecasts of revenues, expenses, income, and earnings-per-share for three years into the future. After recording their forecasts for one company, subjects were provided with feedback on the quality of their decisions by showing them the "accurate" (monte carlo) values, prior to forecasting for the next company. After the first and last trials, subjects were asked to rate their confidence in their decisions and their satisfaction with the company reports on 7-point Likert scales.

The experimental procedure as described was designed to capture the procedural knowledge gained by the subjects as they were repeatedly exposed to a particular method of data display. Our interest was in detecting changes in decision quality over time for each of the three experimental groups. At the end of the study, in order to explore possible differences in declarative knowledge among the three groups, an additional set of company data was presented to the subjects; following development of earnings forecasts, the subjects were asked to draw the reports they had read from memory (as a recall test) and to answer 13 multiple choice questions about the data (as a recognition test). It was hoped that the results of these two tests would explain some of the observed differences in decision quality across the three treatment groups.

Results

Summary statistics for decision quality measures for each of the three experimental groups are shown in Table 1a and 1b. Smaller scores correspond to better decision quality, since decision quality was measured in terms of the average percent forecast error. Because the homogeneity of variance assumption necessary for ANOVA procedures was not met, the raw values for decision quality were converted to a logarithmic scale and all subsequent analyses performed on these transformed scores. Figure 1 shows plots of the mean decision quality scores for each group over the five experimental trials. Simple observation of the means suggests that minimal improvement in performance occurred in the tabular group; there was a slight "learning curve" in the nonstandard graphics treatment and a steeper learning curve in the standard graphics group.

We began the data analysis by performing a multivariate analysis of variance (MANOVA) for repeated measures (Table 2). Bartlett's test indicated that the assumption of sphericity in the variance and covariance matrix was satisfied (Chi square=2.05, $p=.92$). No significant differences among the treatment groups were detected, and the test for a group by trial interaction was likewise nonsignificant. A significant trial effect was observed, however, indicating a change in decision quality over time for all 3 groups combined. These MANOVA results must be interpreted with caution for several reasons. First of all, the tests for group effects were performed on the *average* decision quality across the 5 trials for each group; since all trials were equally weighted, differences in later trials of the experiment might not have been detected by the F test. Second, the repeated measures MANOVA presumes independence of trials, and the use of feedback between trials led to violation of this assumption; again, the model is unable to detect learning effects. Finally, there is the possibility of insufficient power; to be included in the analysis each subject must have made 3 forecasts of EPS for each of the 5 trials. Although 75 subjects participated in the study, missing data points resulted in the loss of degrees of freedom for the MANOVA F test.

In order to detect changes in group performance over time, univariate F tests on decision quality were conducted for each of the 5 trials (see Table 3). Significant differences among the three treatment groups were observed in the fifth trial. A posteriori contrasts using Scheffe's method indicated statistically significant differences between the tabular and standard graphics groups ($T=2.81$, $p=.006$), and between the standard and nonstandard graphics groups ($T=2.00$, $P=.049$). Although graphs were no more effective than tables at the beginning of the experiment, a performance advantage for graphs emerged following practice. Paired (within subjects) t-tests were used to identify improvements in performance within each group over the experimental trials (see Table 4). No learning is evident in the tabular group; some indication of learning is evident in the nonstandard graphics group; and there is clear indication of performance improvement in the standard graphics group.

Along with changes in decision quality, differences in report satisfaction and decision confidence were observed as the experiment progressed. At the end of the first trial, the tabular group was significantly more confident in their decisions than the two graphical treatments. Satisfaction with the reports was also higher in the tabular group than in the other two groups following the first trial (although this difference was not significant). However, by the end of the experiment, no significant differences in ratings of satisfaction or confidence across the three groups were apparent. Thus, the greater confidence associated with using traditional spreadsheets was no longer evident following practice in the forecasting task. Ratings in the tabular group tended to decline over time. Ratings in the standard and nonstandard graphical groups tended to improve over time although the improvements were not significant. The increased confidence in the graphical treatments suggests that subjects felt they improved in task performance (i.e., they learned).

Responses to the 13-item recognition test and the free recall test were next examined for possible indication of differences in declarative knowledge acquisition across the three treatment groups. The recognition test was similar to verbal "interpretation accuracy" measures used in prior studies comparing tables with graphs and contained four types of questions distributed as follows:

(a) 3 questions about trends in the earnings data over time
(b) 4 questions about variable relationships within a chart

Table 1a

Means (and Standard Deviations) for Decision Quality
across 5 Trials for 3 Treatment Groups
[Raw Scores]

	Trial 1 Mean (SD)	*Trial 2 Mean (SD)*	*Trial 3 Mean (SD)*	*Trial 4 Mean (SD)*	*Trial 5 Mean (SD)*	*Total Mean*
Tabular	22.53 (21.5)	14.65 (10.2)	21.82 (28.9)	17.71 (21.6)	17.46 (21.0)	18.83
Standard Graphics	19.57 (17.7)	23.27 (32.9)	15.69 (25.8)	14.24 (17.7)	9.28 (8.83)	16.41
Nonstandard Graphics	24.41 (23.3)	18.31 (19.2)	17.89 (15.8)	13.50 (10.5)	16.01 (19.8)	18.02
Total	22.17	18.74	18.47	15.15	14.25	

Table 1b

Means (and Standard Deviations) for Decision Quality
across 5 Trials for 3 Treatment Groups
[Transformed Scores]

	Trial 1 Mean (SD)	*Trial 2 Mean (SD)*	*Trial 3 Mean (SD)*	*Trial 4 Mean (SD)*	*Trial 5 Mean (SD)*	*Total Mean*
Tabular	1.19 (.41)	1.07 (.31)	1.09 (.45)	1.09 (.35)	1.11 (.32)	1.11
Standard Graphics	1.14 (.37)	1.11 (.48)	.95 (.40)	.98 (.37)	.82 (.36)	1.00
Nonstandard Graphics	1.26 (.32)	1.11 (.39)	1.08 (.40)	1.03 (.31)	1.03 (.38)	1.10
Total	1.20	1.10	1.04	1.03	.99	

Table 2

MANOVA Results for Tests of
Treatment Group, Trial, and Interaction Effects on Decision Quality

Source of Variation	*SS*	*DF*	*MS*	*F*	*Sig. of F*
Within Groups	10.23	51	.201		
Constant	291.77	1	291.77	1455	
Group	.696	2	.348	1.74	.186

R squared = .064

Source of Variation	*Wilks Lambda*	*Hypoth. DF*	*Error DF*	*F*	*Sig.*
Trial	.756	4	48	3.87	.008*
Group by Trial	.906	8	96	.607	.770

(c) 2 questions about variable relationships across charts
(d) 4 questions about specific point values within a chart

Scores for each of these 4 categories, as well as a global score, were computed for each subject. No significant differences were observed for the 3 treatment groups on any of these measures. The results lend little insight into possible reasons for the observed differences in decision quality among the three groups at the end of the experiment.

Analysis of the free recall results were more interesting than those of the recognition test. The recall test was designed to capture spatial declarative knowledge. Subjects were asked to "recreate the financial report which you [just] read. Draw the report from memory . . . Put as much detail in your drawing as you possibly can, including labeling, headings, and numbers." In general, the graphical groups were better able to reconstruct a report than the tabular group. They were more likely to put labels or headings on the report and, in the standard graphics condition, were more likely to put specific numbers within the report. Over 30% of subjects in each graphical treatment were able to put tick marks or exact numbers on the axis scales of their reports, and the majority also attempted to draw a legend for the graph. These findings tend to contradict Vernon's (1946) hypothesis that people attend more to the "raw picture" of a graph than to the contextual details; rather, it seems that users of graphs can retain a great deal of detailed information—at least in a situation where they have had practice in viewing similar graphs. The results of the free recall may provide some explanation for the observed performance differences in the three treatment groups. We can postulate that graphs—in particular, standardized graphs—facilitate acquisition of a more complete 'picture' of the data which, in turn, encourages accurate forecasting. These conclusions should be regarded as tentative since the recall and recognition tests were performed for exploratory purposes and were not validated measurement instruments.

Discussion

The findings of this study provide support for the notion that learning is important for effective use of graphs as decision aids. Graphs initially provided decision makers with no meaningful advantage over tables. Practice in using reports to make financial forecasts led to gradual improvement in decision quality for the graphical treatments but to no meaningful performance change for the tabular treatment. After five experimental trials, forecast accuracy was significantly better with graphs than with tables.

The noteworthy finding in the current experiment is that the advantages of graphs may not be demonstrated behaviorally unless a learning process takes place. Of course, this may only be true in the case of MBA students with approximately four years of work experience in business. One must always be careful in generalizing the results of laboratory experimentation. Nevertheless, the rather large sample size used in this study, and its reasonable similarity to the population of practicing managers, suggests that graphs are in fact novel to many users and that their effective use in decision tasks will require an adjust-ment process. There seems to exist what Lusk and Kersnick (1979) called a "conditioning bond" toward tables, particularly in standard business reporting, and this appears to take time to break down.

For researchers, a major implication of these results is that the role of learning must be either manipulated or controlled when examining the tables-versus-graphs controversy. The best advice seems to be to employ designs with repeated measures and, where possible, observe decision-related behaviors over a longer time period than the traditional 30 to 40-minute experimental session. The conclusions of the current study certainly would have been quite different had the experiment consisted of only one trial. Although it is unknown from the current study whether performance improvement would have occurred without the use of feedback, the results still suggest that the failure of so many prior studies to detect performance advantages for graphs may have been due to the novelty of the presentation format to the subjects. It follows that we must allow research subjects the opportunity to become familiar with a novel technology before taking measurements and forming conclusions regarding the effectiveness of that technology.

An additional issue for researchers to consider relates to the measurement of knowledge acquisition by users of graphical material. The usual approach has been to develop multiple choice and true/false questions; the total number of correct items is then calculated and mean performance then compared for tabular and graphical treatment groups. Many studies have showed little support for graphics using this type of measure. Different results might be found if the items were grouped into categories according to content, such as "recognition of point values," "detection of trends," or "identification of relationships among variables." In other words, the researcher should not assume that a custom-designed instrument is unidimensional, capturing a construct loosely labelled "interpretation accuracy." Beyond this problem, the researcher of presentation modes should also be concerned with the fundamental nature of instruments used to assess declarative knowledge. Multiple choice or fill-in questions are highly dependent on verbal material. Yet graphs are, by their nature, spatial. Asking subjects to draw what they recall, or to respond to visual patterns rather than to words, may be a more appropriate method for assessing the impact of graphs on recall and recognition of information. The difficulties of developing and validating this type of measure are obvious, but the point is we must begin to explore more meaningful methods of assessing the impact of graphs on declarative knowledge acquisition.

Further research effort is needed to determine the precise nature of the learning curve associated with graphs, the cognitive skills required to interpret graphical output, and techniques which might be used to hasten the learning process. The current experiment demonstrated that the knowledge necessary to use graphs in decision making can be acquired with practice. However, within the confines of one study we cannot establish *how long* it takes to adjust to graphics. Also, the shape of the performance curves for the three groups is unknown for more distant points into the future (i.e., beyond trial 5); the curves for the graphical treatments may decline, flatten out, or even merge together. Subjects with nonstandard graphs exhibited some learning during the experiment. Perhaps they would have improved if a greater number of experimental trials had been given. The long run nature of the performance curves for each of the 3 groups would be very interesting to observe. In short, although repeated measures were used, the experiment is still limiting in that it provides us with only a "snapshot" of the behavior we are seeking to understand.

A second issue concerns the cognitive process experienced by users as they view graphical data. This experiment only considered observable behavior, as evidenced in task performance. Subsequent research must study what is learned in addition to how performance changes. That is, researchers must closely examine what happens cognitively as a decision maker "learns" to use graphical tools. Once cognitive requirements for using graphs are understood, techniques might then be developed—that go beyond mere practice and feedback—for improving accurate reading of graphs and their incorporation into the decision process.

In addition to the role of learning, the results of this study suggest the importance of using standards when constructing the axis scales of graphs. Although violation of scaling standards did not lower decision quality beyond the level obtained with traditional spreadsheets, performance in the task was consistently better with standard then with nonstandard scaling. Furthermore, although the users of nonstandard graphs were able to exhibit some degree of learning, they were not able to adequately compensate for scaling distortions within the time span of the experiment. The results imply that guidelines for graph construction should indeed be followed when graphs are developed for decision aiding purposes.

Further research might examine the importance of guidelines for scaling and other graphical components beyond those considered in this particular study. Guidelines similar to those found in scaling are available for the color, size, and shading components of graphs. The effect of violation of proposed standards on perceptual distortion and quality of decision making is in need of empirical study. From a practical standpoint, it seems important to conduct this kind of research within a learning environment. If users can adapt to poor graph design, then the costly process of adding default procedures to software, or training users in good design, may be avoidable.

As a final point, the current study suggests that financial reporting may be an appropriate application area for the use of graphics. Income statements and historical earnings data displayed in graphical format improved forecast accuracy over that with traditional spreadsheets. Considerable work is needed, however, to understand the role of learning in use of graphs and to develop standards for presenting accounting information in graphical form. Further work in graphs as a presentation method for standard accounting information appears to be a worthwhile pursuit.

Table 3

Summary of ANOVA Results for Tests of a Treatment Group Effect on Decision Quality for 5 Experimental Trials

Decision Quality	*DF*	*F*	*Sig. of F*
Trial 1	2, 59	.494	.61
Trial 2	2, 66	.085	.92
Trial 3	2, 68	.856	.43
Trial 4	2, 67	.658	.52
Trial 5	2, 70	4.15	.02*

Table 4

Paired t-tests Within Each Treatment Group

Group	*Hypothesis*	*N*	*DF*	*t*	*Sig. of t*
Tabular	PERF1 > PERF3	20	19	2.27	.13
	PERF2 > PERF4	21	20	−.46	-
	PERF3 > PERF5	23	22	−.11	-
	PERF1 > PERF5	20	19	.30	.38
Standard Graphics	PERF1 > PERF3	21	20	3.08	.025*
	PERF2 > PERF4	21	20	1.35	.097
	PERF3 > PERF5	22	21	1.14	.134
	PERF1 > PERF5	20	19	2.68	.007*
Nonstandard Graphics	PERF1 > PERF3	20	19	1.61	.069
	PERF2 > PERF4	23	22	.84	.275
	PERF3 > PERF5	23	22	.93	.182
	PERF1 > PERF5	20	19	2.62	.008*

Note: "PERF" refers to the performance measure, i.e., decision quality. The number following PERF refers to a particular experimental trial.

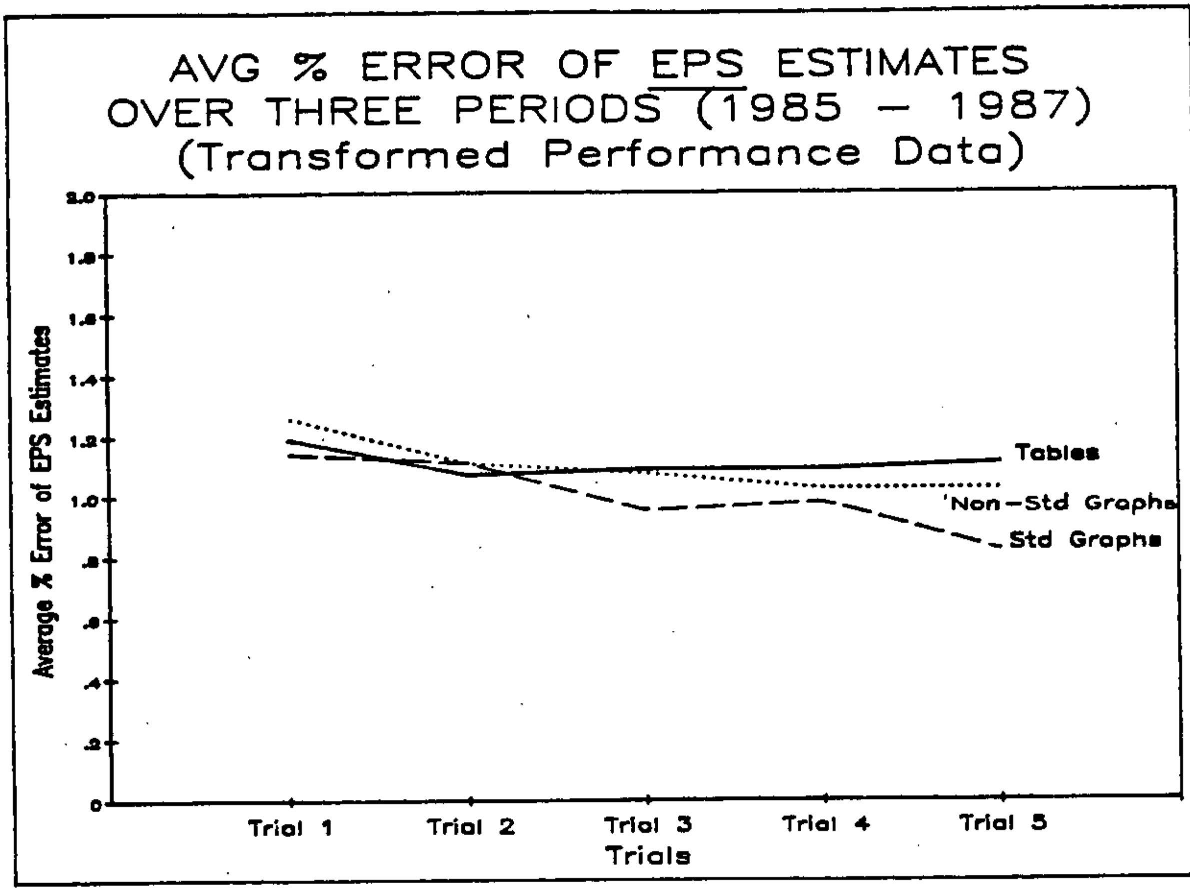

Figure 1

Average % Error of *EPS* Estimates
Over Three Periods (1985–1987)

REFERENCES

Anderson, J.R. "Cognitive Psychology and Its Implications", San Francisco: W.H. Freeman, 1980.

Benjamin, J.J., & Strawser, R.H. "The Publication of Forecasts: An Experiment", *Abacus*, 1974, pp. 138–146.

Bettman, J.R., & Zins, M.A. "Information Format and Choice Task Effects in Decision Making", *Journal of Consumer Research*, Volume 11, 1979, pp. 141–153.

Brandon, C.H., & Jarrett, J.E. "Experimenting With Students' Ability to Forecast", *The Accounting Review*, Volume 52, No. 3, 1977, pp. 697–704.

Brandon, C.H., & Jarrett, J.E. "Revising Earnings Per Share Forecasts: An empirical test", *Management Science*, 1979, Volume 25, No. 3, pp. 211–220.

Brinton, W.C. "Graphic Methods of Presenting Facts", New York: Engineering Magazine, 1916.

Cox, D.R. "Some Remarks on the Role in Statistics of Graphical Methods", *Applied Statistics*, Volume 27, 1978, pp. 4–9.

DeSanctis, G. "Computer Graphics as Decision Aids: Directions for Research", *Decision Sciences*, Volume 15, No. 4, 1984, pp. 463–487.

DeSanctis, G., and Dickson, G.W. "Computer Graphics as Decision Support Tools for Data Interpretation and Trend Spotting", *Proceedings of the Eighteenth Annual Hawaii International Conference on System Sciences*, 1985, pp. 557–562.

Graham, J.L. "Illusory Trends in the Observation of Bar Graphs", *Journal of Experimental Psychology*, Volume 20, 1937, pp. 597–608.

Ives, G. "Graphical User Interfaces for Business Information Systems", *MIS Quarterly*, 1982, Special Issue, 15-47.

Jarett, I.M. "Computer Graphics and Reporting Financial Data", New York: John Wiley and Sons, 1983.

Jarvenpaa, S., Dickson G., and DeSanctis, G., "Methodological Issues in Experimental IS Research; Experiences and Recommendations", *MIS Quarterly*, Volume 9, No. 2, 1985, pp. 141–156.

Kimble, G.A. The definition of learning. In Cummings, L.L., and Scott Jr., W.E. (Eds.), "Readings in Organizational Behavior and Human Performance", Homewood, Illinois: Richard D. Irwin, Inc., 1969, pp. 93–101.

Lusk, E.J., and Kersnick, M. "Effects of Cognitive Style and Report Format on Task Performance: The MIS Design Consequences", *Management Science*, Volume 25, 1979, pp. 787–798.

MacDonald-Ross, M. "How Numbers are Shown", *AV Communication Review*, Volume 25, No. 4, 1977, pp. 359–409.

Mason, R.O., and Mitroff, I.I. "A Program for Research on Management Information Systems. *Management Science*, Volume 19, 1973, pp. 475–487.

Moriarity, S. "Communicating Financial Information Through Multidimensional Graphics", *Journal of Accounting Research*, Volume 17, No. 1, 1979, pp. 205–224.

Nawrocki, L.H. "Alphanumeric Versus Graphic Displays in a Problem-Solving Task", Arlington, VA: U.S. Army Behavior and Systems Research Laboratory, Technical Research Note 227, September 1972.

Olson, J.M. "Experience and the Improvement of Cartographic Communication", *Cartographic Journal*, Volume 12, 1975, pp. 94–108.

Otley, D.T., and Dias, F.J.B. "Accounting Aggregation and Decision-Making Performance: An Experimental Investigation", *Journal of Accounting Research*, Volume 20, No. 1, 1982, pp. 171–180.

Powers, M., Lashley, C., Sanchez, P., and Shneiderman, B. "An Experimental Comparison on Tabular and Graphic Data Presentation", University of Maryland, working paper, February 1982.

Pratt, J. "Post-Cognitive Structure: Its Determinants and Relationship to Perceived Information Use and Predictive Accuracy. *Journal of Accounting Research*, Volume 20, No. 1, 1982, pp. 189–209.

Vernon, M.D. "Learning from Graphical Material. *British Journal of Psychology*, Volume 36, 1946, pp. 145–158.

Watson, C.J., and Drive, R.W. "The Influence of Computer Graphics on the Recall of Information", *MIS Quarterly*, Volume 7, No. 1, 1983, pp. 45–53.