ELSEVIER

International Journal of Industrial Ergonomics 22 (1998) 439–447

International Journal of Industrial Ergonomics

Do grables enable the extraction of quantitative information better than pure graphs or tables?

Jessica K. Hink[a], Jason K. Eustace[b], Michael S. Wogalter[a,*]

[a] *Ergonomics Program, Department of Psychology, North Carolina State University, Raleigh, NC 27695-7801, USA*
[b] *Business Program, Department of Finance, American University, Washington, DC 20016, USA*

Received 15 September 1996; received in revised form 20 September 1996; accepted 27 January 1997

Abstract

Previous research is equivocal regarding the most effective methods of presenting quantitative information displays. The differences in results may be due to numerous reasons including the display and inquiry type. This study examines several methods of displaying quantitative information (e.g., line graphs, line grables, bar charts, bar grables, tables, pie charts and pie grables) that were factorially crossed with different kinds of data extraction inquiries (i.e., questions about exact numerical quantities, comparisons, and trends). Grables are displays that combines features of graphs and tables including specific numerical information with each graphically presented category. Results showed that tables, bar grables and line grables produced the fewest errors, and line graphs and bar charts produced the fastest responses across question types. Error rates combining the accuracy and time (i.e., errors/s) were lowest for the three grables and table. Results are discussed with respect to prior theoretical work and the potential benefits of hybrid forms of quantitative displays for multiple kinds of data extraction inquiries.

Relevance to industry

Choosing the best method of displaying information is important for effective decision making. This study evaluates seven types of graphical displays to answer three types of inquiries. Results indicate that in general, the most efficient data extraction (fewest errors per unit time) were produced using grable or table displays across question types. The appropriate display fosters better communication of information.

Keywords: Graphical-display; Information display; Quantitative information display; Visual displays; Comprehension

1. Introduction

The information age has brought to bear the ability to access vast amounts of quantitative data. The capability to easily retrieve information creates a problem of how to present quantitative data most effectively so it is quickly understandable to the viewer. The goal of most data reduction and summarization techniques is to enable fast and accurate extraction of various kinds of information about specific quantities, trends, and comparisons.

* Corresponding author.

PII S0169-8141(97)00017-6

The most common way of conveying quantitative information is the narrative form, where trends and comparisons are described in words and the summaries of the numerical data (e.g., means, percentages) are contained within the text. However, information is not always effectively conveyed through the narrative form, especially when reporting more than a few individual pieces of data. The Publication Manual of the American Psychological Association (APA) (1994) suggests that small amounts of data should be narratively presented. However, when there is more than three or four numbers, APA suggests that it should be reported in a non-narrative form.

There are two primary methods for visually displaying quantitative information in a non-narrative form: numerical and spatial. The numerical display, generally known as a table, presents specific (precise) quantities in an alphanumeric form. The spatial display, known as a graph or chart, presents the information in a picture form.

It has been a long held belief that it is easier for humans to perceive and reason about quantitative data if it is in a graphic format (Washburne, 1927; Culbertson and Powers, 1959; Schutz, 1961). However, the belief that graphs enable a quick and accurate extraction of information about specific trends and comparisons does not find strong empirical support. In fact there is much debate regarding the use of tables and graphs (Pinker, 1990).

Some research indicates that specific numerical quantities are easier to extract from tables and that table displays are less likely to produce misleading interpretations than graphs (Ehrenberg, 1975; Tufte, 1983; Pinker, 1990; Coll et al., 1993). Tufte (1983) recommends that tables are more appropriate than graphs for data sets containing less than 20 observations. Henry (1993) points out that tables can be useful in displaying large amounts of complex data, as opposed to sophisticated graphs such as three-dimensional and radar presentations. However, there is research that conflicts with these recommendations. Mahon (1977) and Wainer and Thissen (1988) suggest that graphs are better suited for large amounts of data because visual patterns such as trends are more difficult to discern from multi-celled tables than from well-designed graphs (MacDonald-Ross 1977; Spence and Lewandowsky, 1991). Graphs are also more effective for time-ordered data, especially if it is expected that the reader will be required to recall specific facts (Coll et al., 1993). Spence and Lewandowsky (1992) and Sanderson et al. (1989) found that graphs can produce emergent features allowing for faster, more accurate data interpretation. Graphical displays may also have added advantages due to their aesthetic properties. If one's eye is drawn to look at a graph, then it might be given more thorough analysis and use and the data understood more fully (Tufte, 1983; Henry, 1993).

Henry (1993) states that "good graphics facilitate comparisons ... a graphic that captures [one's] attention may increase their understanding and use of important data". Apparently, the belief that data interpretation can be improved through graphs derives from the notion that graphs facilitate the processes of perception and pattern recognition, allowing available resources to be used in performing cognitive operations involved in integration and inference (Wickens and Andre, 1990a; Bennett and Flack, 1992).

There are many forms of graphic displays. The issue of how to best present graphical data was considered over 200 years ago when Playfair (1786) began to examine how graphs could be used to show trends in data. In the 1920s this problem still existed when a debate arose in the *Journal of American Statistical Association* regarding the merits of bar versus pie charts (Eells, 1926; Croxton, 1927; Croxton and Stryker, 1927; Von Huhn, 1927). Later studies have continued the debate over pie charts and bar charts, but none have produced conclusive evidence to support the superiority of either display across situations (Croxton and Stein, 1932; Peterson and Schramm, 1955; Culbertson and Powers, 1959). The question of whether one type of graph is better than others still exists.

One explanation for the potential advantage of graphs compared to tables is Wicken's (1992) proximity-compatibility principle. This principle states that data integration processes are facilitated by an object-like presentation, or in this case, a graphical format. Object displays are advantageous for two reasons: (1) they foster parallel processing, and (2) they are more likely to allow pattern formation that serves to aid information integration.

Given that graphs have beneficial properties which facilitate understanding of quantitative data, many guidelines have been published on graph construction with recommendations and procedures for their creation (Kruswald, 1975; Schmid, 1983; Tufte, 1983; Hollands and Spence, 1992; Pinker, 1990; Kosslyn, 1994). However, Cleveland and McGill (1984) concluded that the standard method for choosing a "… graph design for data analysis and presentation is largely unscientific". Indeed, there is relatively little empirical research supporting most design principles, and the results are frequently equivocal (Henry, 1993). For example, some research supports the use of bar charts over pie charts; other research finds the opposite (Von Huhn, 1926; Eells, 1926; Croxton, 1927; MacDonald-Ross, 1977; Spence and Lewandowsky, 1991).

These equivocal results might be due to the particular images used and the type of task that observers are asked to perform. Spence and Lewandowsky (1991) found that pie charts are better than bar charts for making comparisons among proportions, but bar charts are better than pie charts when making direct magnitude estimates. They concluded that unless one is trying to transmit precise numerical values to the viewer, tables are inferior to charts and graphs. Simkins and Hastie (1987) examined different kinds of graphs for their ability to foster proportional judgment. They found that the pie chart was not significantly different from a bar chart in terms of accuracy, but participants took significantly longer to make a judgment from the pie chart over the bar chart display. Wickens (1992) concluded that a bar display would degrade performance relative to a line display. Carswell (1992) found that judgments requiring focused attention (e.g., seeking precise numerical values) are better performed with bar charts than line graphs. Coll et al. (1991) expanded on the idea that particular judgments are better for certain graphical forms when they suggested "superiority of one mode [graphical display] or the other must be dependent on situational factors". Thus, the discrepant results might be due to the varied kinds of data acquisition tasks and displays employed.

Powers et al. (1984) also attempted to determine which form of data display is the easiest to comprehend. Powers et al. (1984) found that tables alone as compared to graphs alone increased comprehension. However, when both tables and graphs were provided together, slower but more accurate performance was produced compared to their individual presentation. Since accuracy is generally more important than speed in most situations, Powers et al. (1984) recommended using the display form most familiar and comfortable to targetted users.

Additionally, the same display might be used for multiple purposes (i.e., determining exact quantities, forecasting trends, making comparisons) by the same or different persons. Given that its potential uses may vary, display effectiveness might be enhanced if aspects of both graphs and tables were combined into a single form. This display that we call the *grable* combines features of graphs and tables, which might accommodate a wider variety task goals than either graphs or tables by themselves. The additional material in grables might clutter the display hindering data extraction and reducing performance (Tufte, 1983).

The present experiment evaluates seven display types: three conventional forms of graphs (line, bar, and pie), three forms of graph-table combinations (line grable, bar grable, and pie grable), and the table form in three types of information extraction tasks (determining numerical values, analyzing trends, and making comparisons). It is predicted that grables, due to the combined nature of both tables and graphs, will allow for faster and more accurate performance than either graphs or tables alone across inquiry type.

2. Method

2.1. Participants

Participants were 63 undergraduates between the ages of 18 and 36 years of age (67% female) from North Carolina State University who fulfilled a course requirement.

2.2. Materials and design

Tables and graphs were produced by Microsoft Excel 5.0 and were laser printed onto 21.6 cm ×

28 cm (8.5×11 in) white paper in landscape orientation.

A total of 441 sheets of tables and/or graphs were produced based on 21 different scenarios, 3 question types, and 7 display types. Scenarios covered a variety of quantifiable situations including: stock prices, number of mountain bikes sold during a 2 year period, weight loss methods, types of student housing, movie ratings, and miles two salespersons traveled. An example of one of the scenarios in each of the seven presentation methods is shown in reduced form in Fig. 1.

The three question types were: (a) *numerical*, e.g., what was the price of stock 2 during week 5?, (b) *trends*, e.g., if you bought stock 1 during the first week and sold it during the fourth week, would you have made any money?, and (c) *comparisons*, e.g., which stock was less expensive – stock 2 during the third week or stock 1 during the fourth week?

The seven presentation methods were (a) line graph, (b) line grable (line graph with adjacent numbers), (c) bar graph (vertically oriented), (d) bar grable (vertical bar graph with adjacent numbers), (e) pie chart, (f) pie grable (pie chart with adjacent numbers), and (g) table.

All alphanumeric characters were printed in Times font. Axis labels were 12 point. The data labels in the grables were 10 point. All questions were 18 point.

2.3. Procedure

Each participant was provided with a packet of 21 stimulus sheets, a response sheet, and blank paper. Every participant viewed all 21 scenarios which were balanced through the seven display types and three question types across participants. Participants viewed a number, trend, and comparison question for each of the seven display types.

Participants were instructed to answer the question located on the bottom of each display as quickly and as accurately as possible, and to perform any work necessary to formulate their answer on a set of available blank paper before writing their final response. Time was recorded from the turn of the each stimulus page to the turn of the next page. Changing answers once written, was not permitted.

3. Results

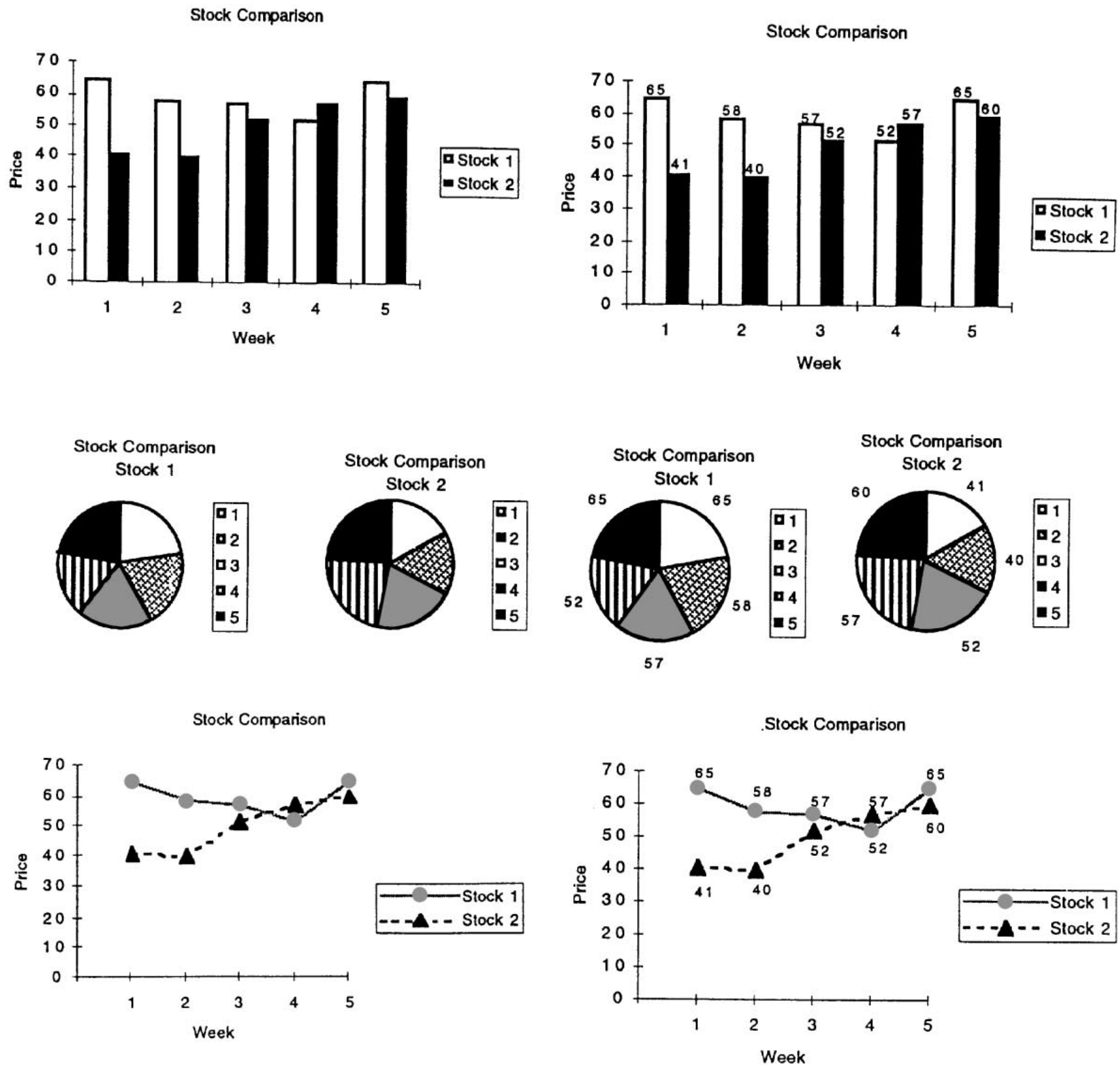
Separate 3 (display type) $\times$ 7 (question type) repeated-measure analyses of variance (ANOVAs) were performed on the accuracy and time data. The error rate in errors per second was calculated using the means from the response time and accuracy data.

3.1. Accuracy

Table 1 shows the proportion correct means. The ANOVA indicated a significant main effect of question type, $F(2, 124) = 3.85$, $p < 0.03$. Comparisons among the means using Tukey's Honestly Significant Difference (HSD) test ($p < 0.05$) showed that accuracy was significantly higher for the comparison questions compared to the number questions. Accuracy on the trend questions was intermediate and not significantly different from the other two question types.

The ANOVA showed a significant main effect of display type, $F(6, 372) = 55.26$, $p < 0.0001$. Comparisons among the means using the Tukey's HSD test indicated that the four displays with numbers (line grables, bar grables, pie grables, and tables) were not significantly different from one another, but all produced significantly more accurate responses than the remaining displays. Bar charts and line graphs did not differ but both produced significantly greater accuracy than pie charts.

The ANOVA also showed a significant interaction of question type and display type, $F(12, 744) = 5.51$, $p < 0.0001$. Comparisons were made using simple effects analyses and contrasts among pairs of means ($p < 0.05$). Examination of Table 1 as well as the tests of significance show a pattern that is consistent with that already described for the display-type main effect (i.e., the displays with numbers did not differ but they all produced greater accuracy than displays without numbers, with the pie chart being the least accurate). However, there were a few exceptions: (a) for the trend questions, the table produced significantly more accurate responses than the line grable; and (b) for the comparison questions, the bar grable produced significantly more accurate responses than the line grable and the table.



Stock Comparison

		Stock 1	Stock 2
	1	$65	$41
	2	$58	$40
Week	3	$57	$52
	4	$52	$57
	5	$65	$60

Fig. 1. Example stack comparison scenario in the seven display forms: bar grable, pie chart, pie grable, line graph, line grable, and table (in reduced size).

Table 1
Proportion correct means as a function of question and display type

Display type	Question type: Number	Trend	Comparison	Mean
Line	0.49	0.67	0.68	0.61
Line grable	0.97	0.71	0.76	0.81
Bar	0.62	0.75	0.81	0.72
Bar grable	0.84	0.84	0.90	0.86
Pie	0.05	0.32	0.46	0.28
Pie grable	0.86	0.75	0.87	0.83
Table	0.90	0.87	0.76	0.85
Mean	0.68	0.70	0.75	

Table 2
Response time means (in s) as a function of question and display type

Display type	Question type: Number	Trend	Comparison	Mean
Line	40.49	49.24	50.63	46.79
Line grable	39.37	60.43	57.14	52.31
Bar	36.84	53.06	47.21	45.70
Bar grable	39.87	60.10	50.67	50.21
Pie	52.98	47.37	57.62	52.66
Pie grable	43.98	71.76	61.52	59.09
Table	30.24	59.24	61.90	50.46
Mean	40.54	57.31	55.24	

3.2. Time

Table 2 shows the mean times (in s) for conditions. The ANOVA showed a significant main effect of question type, $F(2, 124) = 41.42$, $p < 0.0001$. Comparisons among the means using the Tukey's HSD test ($p < 0.05$) indicated that responses to number questions were significantly faster than to trend and comparison questions, with the latter two question types not differing.

Additionally, the ANOVA showed a significant main effect of display type, $F(6, 372) = 5.73$, $p < 0.001$. Comparisons among the means showed that the pie grable produced significantly slower response times than all other display types except the pie chart. No other difference was significant.

The ANOVA also showed a significant interaction of question type and display type, $F(12, 744) = 4.30$, $p < 0.0001$. The cell means in Table 2 show a fairly complex pattern in which response time differs as a function of question and display type. The following description outlines the comparisons that showed significant differences ($p < 0.05$). (a) For number questions, the response times reflect a pattern similar to the main effect of display type already described. The pie chart produced significantly slower response times than all other displays except for the pie grable. The pie grable was significantly slower than the table (for which the fastest times were found). (b) For the trend questions, the pie grable produced significantly slower response times than all other displays. The line grable, the bar grable, and the table produced significantly slower response times than the line graph and pie chart. (c) For the comparison questions, the table and pie grable produced significantly slower response times than the line graph and the two bar displays.

3.3. Error rate

The error and time scores appeared to show hints of a speed–accuracy tradeoff. Some of the displays produced fewer errors but took longer to answer. Because of this circumstance, these two measures were combined to produce a single performance measure, error rate (errors per second), that reflected data extraction efficiency. Inferential tests were not performed on the error rates because the raw error scores were frequently equal to zero, which when divided by time, always produced error rates equal to zero. Instead the data in Table 3 were derived from the overall error and time means for these conditions. The table shows that numerical questions produced the highest and comparison questions the lowest error rate. More importantly these data show that in all cases the grables displays produced lower error rates than its corresponding graph type. Performance on the tables was equal to the grables. Pie charts had the highest error rate. For precise numerical quantities, the line grable is best followed by the pie grable, table, and bar grable. For comparisons, the table, bar grable, and pie grable are best. For trends, the bar grable

Table 3
Error rate in errors/s as a function of question and display type

	Question type			
Display type	Number	Trend	Comparison	Mean
Line	0.0126	0.0067	0.0063	0.0083
Line grable	0.0007	0.0048	0.0042	0.0036
Bar	0.0103	0.0047	0.0040	0.0061
Bar grable	0.0040	0.0026	0.0019	0.0028
Pie	0.0179	0.0143	0.0094	0.0136
Pie grable	0.0032	0.0034	0.0021	0.0029
Table	0.0033	0.0022	0.0039	0.0030
Mean	0.0079	0.0052	0.0045	

and pie grable are best. Thus according to the error rates, tables and the grables are the better choice over than the conventional graphics regardless of question type. With respect to the conventional graphs, the lowest error rates were consistently produced by bar graphs followed by line graphs and lastly by pie charts.

4. Discussion

The results of this study show several discernible patterns. The displays with numbers (tables and the three types of grables) produced the most accurate responses across question types as well as the lowest error rates. Grables, which combine the precise quantities of tables with the emergent features of graphs, performed well with respect to mean errors and error rate compared to the corresponding conventional graph types.

The facilitated performance on the number questions by the table and grables is not unexpected because specific numbers are requested and these displays provide them. Conventional graphs do not display exact data values for specific item categories and thus readers must use a remote scale axis to interpolate estimated values. Pie charts which only show relative category size are the most quantitatively indeterminate and showed extremely poor accuracy not only for numerical but also for trend and comparison inquiries. Nevertheless, pie charts had the fastest response times for trend inquiries.

Because tables lack the visual qualities of graphs, it is somewhat surprising that tables produced relatively high accuracy not just for the number questions but also for the trend and the comparison questions. However, for these two types of inquiries, tables produced relatively slow response times. The longer latency for tables in these conditions might reflect the higher cognitive load involved in transforming the numbers to a more usable visual mental representation, but once formed, the questions can be answered accurately.

Two of the combined displays, line grables and bar grables, took longer to answer in some cases than the simpler line graph or bar chart; this is not surprising because these displays contain more information containing both a graphic representation and specific numbers. However, the extra time produced more accurate interpretations than the conventional graphs. If interpretation accuracy is the primary goal, then bar and line grables appear to be the best choice despite indications of a speed–accuracy tradeoff. With respect to data extraction efficiency, the bar grable appears best for comparison and trend determinations, and the line grable appears best for numerical determinations.

The results have implications with respect to existing recommendations and empirical research. Wickens' (1992) proximity compatibility principle would predict that tasks requiring integration of information such as comparison and trend questions, are better served by more integral, object-like displays such as line graphs but not bar graphs. However, the present study found that bar graphs generally produced better performance than line graphs. However, these results support Sanderson et al.'s (1989) finding that bar graphs support task performance better than line displays but the difference in the present study was not always substantial and significant.

The results also have implications for guidelines with respect to display clutter. Tufte's (1983) data-ink ratio guideline predicts that redundant information such as the inclusion of numerical values in graphs (as seen in grables) would degrade performance. Kosslyn (1994) recommends leaving off specific values because they force the reader to perform more work. Wickens and Andre (1990b), however, failed to show an effect of adding numbers to displays. The current study shows a general benefit of having numbers in the displays, but the facilitation depended on the type of inquiry and the performance measure. For example, while pie charts generally produced poor performance, it produced the fastest response times for trend inquiries.

Additional research on the best ways of displaying quantitative information is needed particularly for the hybrid graph-table that we have termed the grable. One area that needs further investigation includes the size and positioning of the alphanumerics in grables. For example, in bar grables should the numbers be placed inside, above, or on the side of the bar? If the numbers are placed inside the bar, then the emergent features that they might form would not be disrupted. At the same time, bars frequently have shading and so contrast could be reduced without white space surrounding the number. Thus the extent and effect of the tradeoffs involved need to be investigated.

The present research suggests that the grable form of quantitative display has potential for communicating information across varied types of inquiry. Systematic research in this area will likely produce display principles that enhance users' ability to extract information which in turn will facilitate decision making.

Acknowledgements

The authors would like to thank David W. Martin for his comments on an earlier version of this article. Portions of this article were presented at the Human Factors and Ergonomics Society 40th Annual meeting in Philadelphia, Pennsylvania (Hink et al., 1996). Requests for reprints should be sent to: Michael S. Wogalter, Department of Psychology, 640 Poe Hall, North Carolina State University, Raleigh, NC 27695-7801 USA.

References

American Psychological Association (APA), 1994. Publication Manual of the American Psychological Association, 4th ed. Washington, DC.

Bennett, B., Flach, J., 1992. Graphical displays: implications for divided attention, focused attention, and problem solving. Human Factors 34, 513–533.

Carswell, C.M., 1992. Choosing specifiers: an evaluation of the basic tasks model of graphical perception. Human Factors 34, 535–554.

Cleveland, W.S., McGill, R., 1984. Graphical perception: theory, experimentation, and application to the development of graphical methods. Journal of the American Statistical Association 79, 531–554.

Coll, R., Thyagarajan, A., Chopra, S., 1991. An experimental study comparing the effectiveness of computer graphics data verses computer tabular data. IEEE Transactions on Systems, Man, and Cybernetics 21, 897–899.

Croxton, F.E., 1927. Further studies in the graphics use of circles and bars II: some additional data. Journal of the American Statistical Association 22, 36–39.

Croxton, F.E., Stein, H., 1932. Graphic comparison by bars, squares, circles and cubes. Journal of the American Statistical Association 27, 54–60.

Croxton, F.E., Stryker, R.E., 1927. Bar charts versus circle diagrams. Journal of the American Statistical Association 22, 473–482.

Culbertson, H.M., Powers, R.D., 1959. A study of graph comprehension difficulty. Audio-visual Communication Review 7, 97–100.

Eells, W.C., 1926. The relative merits of circles and bars for representing component parts. Journal of the American Statistical Association 21, 119–132.

Ehrenberg, , 1975. Data reduction: Analyzing and Interpreting Statistical Data . Wiley, New York.

Henry, G.T., 1993. Using graphical displays for evaluation data. Evaluation Review 17, 60–78.

Hink, J.K., Wogalter, M.S., Eustace, J.K., 1996. Display of quantitative information: are grables better than plain graphs or tables? Proceedings of the Human Factors and Ergonomics Society, 40th Annual Meeting, pp. 1155–1159.

Hollands, J.G., Spence, I., 1992. Judgments of change and proportion in graphical perception. Human Factors 34, 313–334.

Kosslyn, S.M., 1994. Elements of Graph Design. W.H. Freeman, New York.

Kruswald, W.H., 1975. Visions of maps and graphs. In: Kavaliunas, J. (Ed.), AutoCartoII, Proceeding of the International Symposium on Computer Assisted Cartography. US Bureau of the Census and American Congress on Survey and Mapping, Washington DC, pp. 27–36.

MacDonald-Ross, M., 1977. How numbers are shown: a review of research on the presentation of quantitative data in texts. Audio-visual Communication Review 25, 359–407.

Mahon, B.H., 1977. Statistics and decisions: the importance of communication and the power for graphical presentation. Journal of the Royal Statistical Society, Series A 140, 298–320.

Peterson, L.V., Schramm, W., 1955. How accurately are different kinds of graphs read?. Audio-visual Communication Review 2, 178–189.

Pinker, S., 1990. A theory of graph comprehension. In: Freedle, R. (Ed.), Artificial Intelligence and the Future of Testing. Lawrence Erlbaum, New Jersey, pp. 73–126.

Playfair, W., 1786. The Commercial and Political Atlas. London.

Powers, M., Lashley, C., Sanchez, P., Shneiderman, B., 1984. An experimental comparison of tabular and graphic data presentation. International Journal of Man–Machine Studies 20, 545–566.

Sanderson, P.M., Flach, J.M., Buttigieg, M.A., Casey, E.J., 1989. Object displays do not always support better integrated task performance. Human Factors 31, 183–198.

Schmid, C.F., 1983. Statistical Graphics. Wiley, New York.

Schutz, H., 1961. An evaluation of formats for graphic trend displays: Experiment II. Human Factors 3, 237–246.

Simkins, D., Hastie, R., 1987. An information processing analysis of graph perception. Journal of the American Statistical Association 82, 454–465.

Spence, I., Lewandowsky, S., 1991. Displaying proportions and percentages. Applied Cognitive Psychology 5, 61–77.

Tufte, E.R., 1983. The Visual Display of Quantitative Information. Graphics Press, Chesire, CT.

Von Huhn, R., 1927. Further studies in the graphic use of circles and bars: a discussion of Eels experiment. Journal of the American Statistical Association 22, 31–36.

Wainer, H., Thissen, D., 1988. Plotting in the modern world: Statistical packages and good graphics. Chance 1, 10–20.

Washburne, J.N., 1927. An experimental study of various graphic, tabular and textual methods of presenting quantitative materials. Journal of Educational Psychology 18, 361–376, 465–476.

Wickens, C.D., 1992. Engineering Psychology and Human Performance. Harper Collins, New York.

Wickens, C., Andre, A., 1990a. Proximity compatibility and the object display. In Proceedings of the Human Factors Society 32 Annual Meeting. Human Factors Society, Santa Monica, CA, pp. 1335–1339.

Wickens, C., Andre, A., 1990b. Proximity compatibility and display: effects of color, space, and objectness on information integration. Human Factors 32, 61–77.