DOCUMENT RESUME

ED 261 385 CS 209 254

AUTHOR Soderston, Candace
TITLE An Evaluative and Prescriptive Look at Graphics Research.
PUB DATE May 83
NOTE 5p.; In: Proceedings of the International Technical Communication Conference (30th, St. Louis, MO, May 1-4, 1983).
PUB TYPE Information Analyses (070) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Evaluation Methods; *Graphic Arts; *Research; Research Design; Researchers; *Research Methodology; *Research Needs; *Research Problems; Research Reports; Research Skills

ABSTRACT

Graphics research history shows that some topics have been studied heavily while others have been almost entirely neglected. Furthermore, researchers have used many different methods of defining and measuring effects such as legibility and comprehension, and this, together with vagueness in reporting, makes it difficult to compare studies and draw reasonable conclusions. Current research is needed to validate past studies and to bring to light reader reactions. This current research should (1) be user and task specific; (2) study interaction effects and optimum concentrations; (3) employ prolonged periods of time and lengthy passages; (4) study the interaction of familiarity with the effectiveness of a device; (5) focus on the gaps in existing research, such as paper, ink, binding, color, and illustrations; (6) report clearly how the research is conducted, what is measured, and what is found (statistical results); and (7) show which methods measure which attributes most effectively. (Author/DF)

AN EVALUATIVE AND PRESCRIPTIVE LOOK AT GRAPHICS RESEARCH

Candace Soderston
Rensselaer Polytechnic Institute

Recently accepted position:
Associate Information Developer
IBM Corporation
64M, 728
Kingston, NY 12401

Graphics research history shows that there are topics that have been studied heavily and others that have been almost entirely neglected. Furthermore, researchers have used many different methods of defining and measuring effects such as legibility and comprehension, and this, together with vagueness in reporting, makes it difficult to compare studies and draw reasonable conclusions. Current research is needed to validate past studies and to bring to light more of the underlying principles that guide reader reactions. This current research should: 1) be user and task specific; 2) study interaction effects and optimum concentrations; 3) employ prolonged periods of time and lengthy passages; 4) study the interaction of familiarity with the effectiveness of a device; 5) focus on the gaps in existing research (paper, ink, binding, color, illustrations); 6) report clearly how the research is conducted, what is measured, how it is measured, and what is found (statistical results); and 7) show which methods measure which attributes most effectively.

AN EVALUATIVE AND PRESCRIPTIVE LOOK AT GRAPHICS RESEARCH[1]

Certainly one of the most crucial skills required of a technical writer is the ability to choose and employ graphic devices wisely. By graphic devices I mean all non-textual, non-content, non-prose related variables that go into the presentation of a finished document. Therefore, I include typography, layout, illustration, color, paper and binding, all as species within the umbrella genus of graphics.

Literature and support documentation are key factors in the ease of use of a technical product and, in the long run, its successful marketing. Technical writers, to support products and provide customers with quality documentation, must be skilled in handling language, organizing content, and designing layout. The use of graphic devices are of prime importance in organizing content and, of course, in designing layout.

Heretofore, much of our handling of graphic devices and tools has been based upon an intuitive, innate feeling for the quality of a piece of work. While an intuitive sensitivity for the effective use of graphics is valuable in a technical writer, it is not sufficient for training future professionals.

Every profession has a history of development, during which time the general laws and underlying principles governing its subject are brought to light. Our profession has drawn its support from, among other areas, cognitive psychology (memory and information processing), engineering (information theory and human factors), composition research, rhetorical theory, and communication theory.

In order to strengthen our professional foundation, train future communicators, and produce the high quality documentation which will compete successfully in the marketplace, we need to continue to push back the borders of the unknown. At this point, taking stock of the existing graphics research will tell us what questions need to be answered by future research. This scrutiny, showing us some flaws in research methodologies, will allow us to avoid the same mistakes in the future.

Actually, a survey of graphics research unearths an inordinate amount of material. Close examination shows, however, that this research, still in a state of infancy, consists of heavily studied topics interspersed among vast neglected areas. Typography, for example, has been the most extensively studied area: an abundance of research has been done on type faces, type sizes, line lengths, interlinear space, column widths and lengths, widths of margins, justification of margins, headings, and cueing devices (caps, italics, boldness, extensions into the left margin, and underlining). On the other hand, very little work has been done on the relative effectiveness of illustrative materials such as charts, graphs, line drawings, half-tones, and photographs. Similarly, few studies have dealt with the effectiveness of color, either as a cueing device or in illustrations, ink, or paper, and virtually no research has looked at the effects of paper and binding on readability or affective reader preference. To conclude from this situation, future research should focus on the neglected areas to give us the information we need in order to design the most effective technical

documents.

Measurement Methodologies in Graphics Research

Past research has studied the effects of graphic devices on readability, legibility, retention, comprehension, aesthetic preference, and, in a few instances, task performance (see Table 1).

Table 1

Effect of Graphic Techniques on Attributes:
Taxonomy of Measurement Methodologies

ATTRIBUTE	MEASUREMENT METHODOLOGIES	
Readability	Eye Movement range of movement pupil size # involuntary blinks # fixations # regressions	Speed of Reading Reading Accuracy upside down rightside up Cloze Procedures
Legibility	# of words read during a brief exposure & correctly reported (tachistoscopic examination) Speed of Recognition (with accuracy) Recognition of letters in peripheral vision " " " at a distance " " " using focal variator (blurring image)	
Retention	Recall Recognition Rate of decrease in retention with increased length of material	
Comprehension	Multiple Choice Test Fill-in Questionnaire Rater's judgements of comprehensibility of passage	Cloze Procedures Open Recall (Storage & Retrieval) Ratio of core/enrichment content learned to total amount learned
Aesthetic Preference	Analysis of introspections Fill-in Questionnaire (Likert-type scales) Ranking adjective pairs describing enjoyment/helpfulness	

These are the methods that researchers have used in the past to measure the effects of graphic devices on readability, legibility, retention, comprehension, and reader preference.

To interpret the results of these studies for our own applications and uses, we have to know exactly what effect was measured and how it was measured. Unfortunately, many older studies are unclear in this area; future research reporting must detail this information clearly and thoroughly.

The bulk of graphics research is old, having been carried out between 1920 and 1965. Most of it was done by English researchers for the British Civil Service and by American researchers for the U.S. military. We must question whether the average English or military user has enough in common with our targeted populations to allow us to generalize the results from these studies to our applications.

Furthermore, because the research is old, many of the data collection methods are now considered to be outmoded and have been discredited. These methods may have distorted the results and, for this reason, the older studies need current validation before we can rely upon their findings and recommendations. For example, tachistoscopic examination of material in testing legibility has been generally disregarded as outmoded and unreliable (a tachistoscope flashes an image briefly upon a screen, and subjects are then asked to identify what they saw: the greater the number of correct identifications, the more legible the image is said to be). Ulric Neisser states that validity is violated in tachistoscopic examination of material, as we do not perceive or read material in isolation from its surrounding material:

> I believe that important aspects of the normal environment are being ignored in contemporary research paradigms. These aspects are the spatial, temporal, and intermodal continuities of real objects and events ... Tachistoscopic experiments simply do not tap normal perceptual skills, and the term perception cannot be consistently applied to anything that occurs in them.[2]

This means that many of the older studies must be suspect, including the works summarized by Vernon (1931).

Recently, researchers have concluded, for the same reason, that the older methods of measuring eye movements were ineffective in assessing legibility and readability: they violated ecological validity. The older methods of measuring eye movements were so intrusive that they removed the reading experience so far from the natural as to make any generalization doubtful. However, in 1975, Whalley and Fleming introduced a new camera system for recording eye movements. This system allows for a more naturalistic reading environment during testing, and may therefore be used to document the validity of the older findings. However, the theoretical

assumptions behind using eye movements to measure legibility must be examined. Tinker (1963) says that blink rates do not reliably measure the level of reading difficulty (people blink their eyes for many different reasons).

We need, at this point, to examine and explore all of the measurement methodologies (again, see Table 1) and, in the future, use only those that are most sensitive and reliable. Hartley, Fraser, and Burnhill (1975) tested nine different measurement techniques and computed reliability coefficients for each. It is interesting to note that they found that the questionnaire is not a reliable method of assessing readability.

Contradictory findings in Graphics Research

As one might imagine, graphics research history is fraught with contradictory findings (see Table 2). Two prominent English graphics researchers, for instance, disagree on the effects of margin width: Burt (1959) says narrow margins produce visual fatigue while Spencer (1968) says they do not. To compare studies and judge results, we have to know the test audience characteristics, the subject matter, the degree of difficulty and length of the reading passages, and the measurement methodologies. For example, if one researcher measures visual fatigue by the number of eye fixations and the other by the amount of involuntary blinking or the speed of reading, our conclusions could be very different. If we are interested in the speed of reading, we would design our material according to the results of studies that measure the influence of graphic devices on speed of reading.

Interaction Effects and Relationships

The effectiveness of a particular device or technique is based on several things: the task characteristics, the user characteristics, and the subject matter, content, and length of the material (see Table 3).

Table 3

Items Affecting the Results of a Study
• User Characteristics • Task Characteristics • Subject, Content, and Length of Material • Presentation Format (Graphic Device) • Experimental Design • Measurement Methodologies

Additionally, there may be interaction effects between graphic devices, and there may be optimal levels (beyond which the effectivity decreases) for a particular technique. For example, underlining as a cueing device may be differentially effective in the following two situations:

- For secretaries trying to locate items within a text that is 10 pages long.
- For engineers trying to memorize key points within a text that is 5 pages long.

Additionally, the effectiveness of underlining may be influenced by type size and interlinear space. One last point to consider is that underlining may be incrementally more effective up to a certain concentration and decrementally effective (actually hamper performance) after that point. I'm sure we would all agree that underlining every other word in a passage for emphasis would produce information overload, hampering both locating and memorizing tasks.

Older graphics research deals exclusively with one causal variable (graphic device) and one effect (measured). Unfortunately, graphic devices most probably interact with each other, with user characteristics, and with task requirements. For complex interaction and naturalistic studies, which reveal a truer picture, the experimental design and analyses are quite complex and have only been developed within the past 10 to 15 years. Future graphics research, to be useful, must be user and task specific, must

Table 2

Examples of Contradictory Findings in Graphics Research

GRAPHIC DEVICE	EFFECT, FINDING, OR RECOMMENDATION	RESEARCHER
Narrow Margins	Produce more visual fatigue Don't produce more visual fatigue	Burt Spencer
Character Size	10 point recommended 12 point recommended Height:Width ratio of 10:7.5 and Stroke:Width ratio of 1:10	Burt Demillo Soar
Chunked Typography	Less Effective More Effective No Difference	Carver Frase & Schwartz Klare, North
Line Length	19 picas recommended Disputed (no reliable difference)	Demillo, Paterson Hartley, Burnhill
Underlining	Improves performance Does not improve learning	Schutz Katzman & Nyenhuis
Serif or Sans-serif	Sans-serif superior Serif superior	Taylor Burt, Spencer

study interaction effects, and must attempt to determine optimim concentrations (see Table 4).

Table 4

Task, User, and Passage Characteristics Affecting Studies and Affecting Graphic Device Impact

Tasks	
• Identifying	• Memorizing
• Locating	• Analyzing
• Counting	• Drawing
• Recognition and Searching	
• Comparing and Verifying (static or dynamic comparisons)	
• Internal or External Pacing	
Users	
• Learning Characteristics	
• Visualizers or Verbalizers	
• High or Low Spatial Aptitude	
• Grade Level	
• Degree of Familiarity with Critical Parameters	
• Background, Expertise	
Passage	
• Subject Matter & Content	
– Level of Vocabulary	
– Level of Conceptual Difficulty	
– Level of Readability	
• Length or Amount of Text	

Time Concerns

There are four additional concerns to take into account when interpreting results from past studies and when planning future ones:

- Length of the reading task.
- Amount of study time permitted the user.
- Familiarity of the user with the particular device.
- Timing of test deliverance (immediate or delayed post-test).

The effectiveness of a graphic device may depend on any or all of these considerations. To begin with, most graphics research has used very short intervals of time so we don't really know how users are affected by the devices under longer reading situations. Tinker (1955) did go back to see if his results would be verified with longer reading tasks (10 minutes versus 1.75 minutes) and they were, so we do have some reason to believe that past results can be generalized to longer task periods. As most technical documents and manuals are lengthy, future graphics research should use prolonged periods of time.

A graphic device can also be differentially effective depending upon whether the users are given only a certain amount of time in which to study the passage or perform the task, or whether they are allowed to proceed at their own pace. For instance, Dwyer found that for internally paced material, illustrations containing a high amount of realistic detail were most effective. Conversely, for externally paced material, a small amount of realistic detail proved to be most effective.

The effectiveness of a graphic device may also depend upon the user's familiarity with it. For example, if users are unfamiliar with a particular format or device, they may be negatively affected at first, but gain in performance or speed (or whatever effect is being measured) as it becomes more familiar. Therefore, it might be wise to pre-test subjects to innoculate or familiarize them with the form t before testing. This is appropriate for graphics research because we may be more interested in the effectiveness of a device on a long-term user rather than on a novice.

Lastly, whether we are more interested in a device's immediate effects or its effect on, say, retention of material, will determine whether we should administer immediate or delayed post-tests.

Conclusions

Based upon a current assessment of past graphics research, future research is needed. This research should help us to both produce higher quality documentation and to more effectively train future technical communicators.

SUMMARY

Guidelines for Future Research

Future graphics research should:

- be user and task specific.
- study interaction effects and optimum concentrations.
- employ prolonged periods of time and lengthy passages.
- study the interaction of familiarity with the effectiveness of a device.
- focus on the gaps in existing research (paper, ink, binding, color, illustrations).
- report clearly how the study is conducted, what is measured, how it is measured, and what is found (statistical results).
- show which methods measure which attributes most effectively.

---

[1]This article developed, in part, from my participation in a research project conducted at the Rensselaer Polytechnic Institute for IBM under contract #S945358QLSAE56. The research team consisted of Dr. Robert Krull (Research Director), Annette Bradford, A.B.D., Dr. Philip Rubens, Dr. Merrill Whitburn, and myself.

[2]Ulric Neisser, Cognition and Reality: Principles and Implications of Cognitive Psychology (San Francisco: W.H. Freeman and Company, 1976) pp. 34 & 46.