Computer Screen Designs: Viewer Judgments

☐ R. Scott Grabinger

This study examined viewer judgments about the readability and studyability of two sets of computer screens: a set of model displays and a set of real screens copied from CAI programs. The purpose was to identify constructs that could guide the design of computer screens used to display information in computer-assisted instruction, hypermedia, or on-line help applications. It also searched for any relationships among viewer preference and viewer field articulation (field-dependence/independence), conceptual style (relational/analytical), and gender. Findings based on multidimensional scaling techniques confirmed and further defined the existence of evaluative constructs based on visual complexity and organization. No generalizable effects for field articulation, conceptual style, or gender differences were found.

☐ The most common method of presenting written instructional information is through paper publications: journals, magazines, books, newspapers, and so on. While paper publications maintain their popularity, the use of electronic publication is growing rapidly. Electronic publications are usually presented on video display screens, including teletext, videotex, computer-assisted instruction, hypermedia, hypertext, and on-line help systems.

Whether a publication is paper or electronic, its production involves arranging numerous text elements—type, illustrations, white space, and graphic devices—so as to present information in a way that communicates the author's message. The combination of these elements to create a meaningful message is not a simple matter. While aesthetic guidelines exist to help designers create attractive displays (Donahue, 1978; Garrett, 1967; Nelson, 1978; Turnbull & Baird, 1975), aside from Hartley's (1985) work, there are few, if any, empirically based guidelines to help instructional designers combine text elements in ways that facilitate learning (Misanchuk, 1992). Waller (1980) states:

> [W]hile rules for the composition of graphically simple verbal sequences exist in usages, and have been described in grammars, the state of typographic theory is relatively primitive. That is, although rules of some kind exist . . . they have not yet been formally codified. (p. 242)

CAI designers need more than aesthetic guidelines. They need guidelines that are focused on learning, guidelines that will help instructional designers create displays that facilitate the processes of reading and learning—the acquisition, organization, and processing of information by learners.

TABLE 1 □ Research Scenario with Eight Text Element Variables and Associate Values

Element	*Values*	
1. Line length	Short	Long
2. Directive cues	Present	Not present
3. Paragraph indication	Spaced	Indented
4. Status bar	Present	Not present
5. Line spacing	Single	Double
6. Functional areas	Present	Not present
7. Text columns	Single	Double
8. Illustrations	Present	Not present

RESEARCH PROBLEM

Developing empirically based guidelines to combine large numbers of text elements to create meaningful screen displays produces three significant research problems. First, research into each possible combination of text format variables and their interactions is a daunting proposition. For example, a study comparing just two values for each of eight text element variables such as those shown in Table 1 yields 256 unique stimuli (2^8), along with the frightening prospect of interpreting eight-way interactions. It is also unrealistic to think that any designers will limit themselves to just eight elements with two values, so research results based on highly controlled examinations of combinations of text elements have limited generalizability.

Second, in any particular layout, the unique contribution of each text element variable to the overall meaning of the display is likely to be quite small (Grabinger, 1989). There may be no way to ascertain to what degree a single variable contributes to the readability of a display or to the amount a reader learns. For example, paragraph indication (e.g., indented or double-spaced) is a single text element variable. By itself, paragraph indication probably has no measurable effect on learning. However, when combined with other variables to organize a display, to create chunks of ideas, or to indicate a hierarchical structure, it may affect how a learner organizes and processes the text and subsequently affect learning.

Finally, a research approach that studies individual text element variables may shed light on how readers perceive and recognize those text elements, but does not provide information about how the reader comprehends, organizes, and processes the information represented by those elements. Determining the best ways to present elements on a page or screen begins with a look at the processes of perception and reading.

Past research in the area of legibility (Rehe, 1979; Reynolds, 1979; Tinker, 1963) provides standards for displays that gain attention and are perceptible and recognizable. However, there is more to reading than recognizing symbols; readers and students must also organize and integrate (Tinker & McCullough, 1962). Since the focus of a screen designer is on devising a combination of format variables to enhance reading and studying, then a complete set of screen design guidelines must go beyond legibility standards and indicate what designers should do to enhance the processes of organization and integration. Bovy (1981) describes this instructional goal as applying methods which "externally model appropriate cognitive processes, direct the cognitive processes of the learner, or allow the learner to activate appropriate methods independently" (p. 208). Therefore, empirically based guidelines developed over the long term must suggest ways to create screens that activate the appropriate processes by the users.

RESEARCH STRATEGY

If activating appropriate processes (i.e., reading and perception) is one of the objectives of an effective set of screen design guidelines, then there are three basic events that a designer can try to enhance: (1) getting the learner's attention, (2) helping the learner find and organize pertinent information, and (3) integrating that information into the learner's knowledge structure. One strategy is to focus on users of potential instructional displays and find out how they judge the displays in terms of readability and "studyability." Studyability refers to the ease with which a user can examine and learn from a screen of information. If evidence of an underlying rationale for such judgments is found, then it can be interpreted to provide information about designing screens. Grabinger (1984, 1987; Grabinger & Amedeo, 1988) found consistent rationales for the judgments expressed by viewers for model screens. Using multidimensional scaling and factor analyses techniques, these studies found that groups of viewers judged sets of model screens as readable or studyable using three evaluative criteria or constructs: structure, organization, and spaciousness/simplicity.

Structure. The most salient criterion, structure, refers to the viewers' preference for screens that appear designed in a way that reflects the content of the subject matter. Structured screens use directive cues for emphasis, headings as organizers, and graphic devices to separate the information into chunks and to indicate important material. Structure is a refined stage of organization.

Organization. Screens that appear to have a coherent arrangement of all the major elements are perceived as organized. Organized screens are divided into functional areas: text, graphics, title/status bar, and navigation control panel. Organization is a precursor to structured screens. Screens that are structured must be organized, but organized screens are not always structured in a way that represents the nature of the content.

Spaciousness. When structure and organization are not present, viewers of computer displays look for spaciousness and simplicity. Some viewers prefer screens that have a lot of white space, little text, few buttons, and few headings. This seems the opposite of structure until one realizes that viewers have priorities when examining screens. If their primary criteria for judging a screen as acceptable for reading and studying are not present, then they fall back to secondary or tertiary criteria.

These earlier studies were conducted using model screens, or screens composed of X's, O's, and I's—the Twyman notation system (Twyman, 1981). Model screens were used to control for content effects and to facilitate interpretation of the data analyses. However, there is still no evidence that the constructs identified with model screens hold up when viewers judge the readability and studyability of a selection of real screens from actual programs (Ross & Morrison, 1989). Therefore, this study expands on the earlier investigations by comparing the findings for a group of model screens with a group of real screens. Using real screens takes a further step in investigating whether the same constructs used by viewers in the model screens emerge in judgments about real screens.

The three dimensions found in the earlier studies indicate that users form judgments about screens based on more than one reason, depending on the importance they assign to their own rationales. However, there was no evidence to indicate why some people prefer structure more than spaciousness or vice versa. It would be useful in the development of guidelines to determine whether groups of viewers have any descriptive characteristics based on common learner analysis tools: conceptual style, field articulation, and gender. Therefore, this study assessed preferences for screen designs as a function of individual differences in field articulation, conceptual style, and gender. The specific goals of this research were as follows:

1. Compare the effects of screen models versus actual screens from programs on viewer preferences for readability and studyability;
2. Investigate the relationship between viewers' preferences and their field articulation, conceptual style, and gender.

ANALYSIS

Analysis Technique

Preference data from paired-comparison tasks were analyzed using multidimensional scaling (MDS) procedures (Hair, Anderson, Tathman, & Grablowsky, 1979; Kruskal & Wish, 1978; Schiffman, Reynolds, & Young, 1981) to identify evaluative dimensions used by subjects while expressing their judgments. MDS is a set of statistical analysis techniques that help reduce and systematize data in areas where organizing concepts and underlying dimensions are not well developed (Schiffman et al., 1981). Therefore, the primary reason for the choice of MDS as an analytical tool is that the dimensions representing reader perceptions about text displays are unknown. MDS does not "require a priori knowledge of the attributes of the stimuli to be scaled. Rather . . . MDS provides a space that reveals the dimensions relevant to the subjects" (Schiffman et al., 1981, p. 3). In other words, MDS is an investigative tool useful for discovering new ways to examine and explain phenomena.

MDS techniques assume that participants report their perceptions of stimuli unidimensionally. Asking the reader for a combined judgment about the readability and studyability of a particular screen produces a unidimensional measure with an ordinal ranking of stimuli. For example, given a selection of 5 screens, a subject will first examine each possible pair: screen 5 versus screen 1, screen 4 versus screen 1, and so on. After selecting the most studyable and readable screen from each pair, the subject may end up with a ranking of 3, 5, 2, 4, 1, screen 3 being the most readable and studyable and screen 1 being the least readable and studyable.

After the rankings for each subject are calculated, the rankings are statistically separated by multidimensional scaling analysis programs to represent distances among the stimuli along a series of one or more dimensions. The distances among the stimuli (screens) indicate how similar and how different the stimuli are from one another as perceived by the group of subjects. These distances (differences and similarities) among the stimuli are represented by points in a multidimensional space. The MDS data analysis provides a set of two-dimensional maps with the stimuli located as points on the map to aid in interpretation (see Figure 6 for an example). The dimensions that compose the axes of the spatial configuration of the stimuli represent the criteria used by the subjects to judge the stimuli. For example, the X axis may represent structure while the Y axis represents organization. Unfortunately, MDS does not provide definitions of the dimensions that result. The definitions of the dimensions are interpreted by visually examining the characteristics of the stimuli that lie along the dimension. The purpose of MDS is to determine the optimal number of dimensions (criteria) that account for the position of objects relative to each other.

Data Collection and Analysis

Subjects viewed two computer screens and performed the paired-comparison task by rating the studyability and readability of the samples shown on the screens. Following the rating task, the multidimensional scaling (MDS) analysis proceeded in two stages. The first stage was to determine from the data the number of dimensions the participants used to evaluate the readability and studyability of the screens. In the second stage, the dimensions that represented the participants' judgments were interpreted by examining the attributes of the screens (stimuli) along the dimensions. The statistical analysis proceeded in the following steps:

Step 1: For each subject, a matrix was created to represent the ranking the subject gave to each screen examined during the paired-comparison test (see Figure 1). The paired-comparison test required two responses. First, subjects indicated on a response form (Figure 2) which screen they thought was most readable and studyable by circling L for left and R for right. Next, subjects indicated how much more readable and studyable their choice was by placing a vertical mark through a 100-millimeter horizontal line on the response form. The marks were scored by measuring the distance of each mark from the left end of the line. The distance in millimeters was then transcribed to a scale of 50 to 100. This value (x) was placed in the column of the matrix representing that stimulus pair, while $100 - x$ was placed in the row of the matrix representing that stimulus pair. The columns of the matrix thus represented the stimulus the subject preferred most, while the rows of the matrix

FIGURE 1 ☐ Sample Participant Matrix and Utilities

Screens	1	2	3	4	5
1		90			75
2	10			65	95
3					
4		35			
5	25	05			

Utilities	4	2	5	3	1

represented the stimulus the subject preferred least.

For example, suppose that a subject views screen 1 on the left and screen 5 on the right. The subject then circles the letter L to indicate that she or he thinks that the lert screen (1) is more readable and studyable than the right screen (5), and places a line midway on the scale to represent how much more readable and studyable the choice was. This line is measured as 50 millimeters (x) and transcribed to a value of 75 [$x + (100 - x)/2)$]. As shown in Figure 1, the value 75 is placed in the column and row of the matrix (upper right) representing the stimulus pair and the value 25 (i.e., 100 − 75) is placed in the row and column (lower left) representing the stimulus pair.

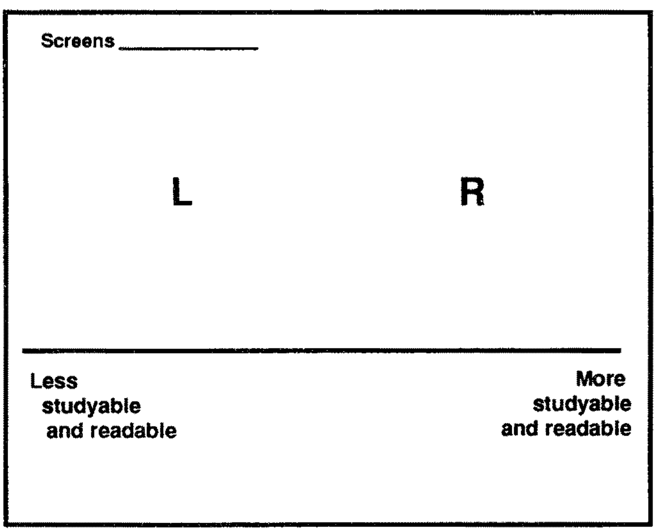
The end result of this step is the creation of a matrix for each subject composed of the ratings assigned to each stimulus pair viewed by the subject. Note that Figure 1 is an incomplete matrix. In the actual studies, an incomplete matrix design (Spence & Domoney, 1974) was used to prevent subject fatigue from examining hundreds of comparisons. A complete matrix design represents a comparison of each possible pair by every subject. However, this is impractical, so an incomplete matrix design has each subject view a subset of the complete set of pairs, creating a partially completed matrix.

Step 2: Torgerson's (1958) procedures for standardizing incomplete matrices were used to transform the raw data into utilities (rankings) for the subsequent multidimensional unfolding. Utilities for each participant are standardized scores representing the rankings of the screens judged by the viewers.

Step 3: The utilities, or standardized rankings, created in Step 2 were analyzed by multidimensional scaling techniques using the SPSS-X ALSCAL procedure (SPSS, 1988).

Step 4: The MDS dimensions were interpreted and defined through three means: First, a visual analysis was conducted by examining and comparing the attributes of the stimuli as they were positioned along the resulting

FIGURE 2 ☐ Sample Response Form for Collecting Data



MDS dimensions. Second, the overall preferences for the screens were calculated to help define which screens were preferred. Third, results of regressing 13 adjectives on the MDS coordinates were used to help define the meanings of the dimensions.

STUDY 1

Method

Participants

Participants were 94 graduate and undergraduate students (41 male, 53 female) from three major universities in Pennsylvania, New York, and Colorado. Subjects' ages ranged from the early 20s to over 60. Subjects were members of a variety of instructional technology classes and received credit for participating in the study. (Table 2 provides a summary of participant characteristics.) Although subjects were members of instructional technology classes, they were not familiar with screen design concepts.

Pretests

Subjects were given tests on embedded figures and conceptual styles before the treatment. The Hidden Figures Test (Educational Testing Service, 1975) is designed to test an individual's ability to locate a simple figure within a larger and more complex figure which has been organized so as to obscure or embed the target figure. The larger the subject's score, the more able the person is to disembed the simple figure, and thus the more field-independent or analytical the person is. The Conceptual Styles Inventory (Kagen, Moss, & Siegel, 1963) is a 30-item graphic test to measure a respondent's predilection to think in a relational or analytical manner. Only Part 1 of the inventory (15 items) was used. These two tests were chosen to investigate whether analytical viewers react to the screens, and perhaps to individual text elements, differently from less analytical viewers.

Primary Task and Materials

Study 1 collected viewer judgments using the paired-comparison task of model screens, i.e., screens that control for content effects through the use of Twyman's notation system (Twyman, 1981). (Samples of all screens are shown in Figure 3.) Eight primary text elements, each with two values, were used in the design of the screens to aid in the interpretation of the MDS results. Table 3 presents the text element configurations of the 32 screens.

TABLE 2 ☐ Participant Characteristics

	STUDY 1 Model Screen Participants (N = 94)		*STUDY 2 Real Screen Participants (N = 33)*	
Variables	*N*	*%*	*N*	*%*
Gender				
Female	53	83	7	70
Male	41	17	23	21
Not specified	0	0	3	9
School				
Colorado	44	47	33	100
New York	32	34		
Pennsylvania	18	19		
Age				
0 to 28	37	39	5	15
28 to 39	30	32	8	24
40 +	25	27	17	52
Not specified	2	2	3	9

An X in the table indicates that that variable value was present in the design of the model text. For example, examine screen 1 in Figure 3 (screen numbers are in boxes at the lower left side) and note that it contains defined functional areas, directive cues, spaced paragraph indication, single line spacing within text, a model illustration, no status bar, and two columns of text. Descriptions of the eight primary text elements follow.

TABLE 3 ☐ Profile of Variables Used to Construct Model Screens

	FA		DC		PI		HD		LS		IL		SB		TC	
	Y	N	Y	N	Sp	In	Y	N	Ss	Ds	Y	N	Y	N	1	2
Screen 01	x		x		x		x		x		x			x		x
Screen 02	x		x		x		x			x		x	x			x
Screen 03	x		x		x			x	x			x	x			x
Screen 04	x		x		x			x		x	x			x		x
Screen 05	x		x			x	x		x		x			x	x	
Screen 06	x		x			x	x			x		x	x		x	
Screen 07	x		x			x		x	x			x	x		x	
Screen 08	x		x			x		x		x	x			x	x	
Screen 09	x			x	x		x		x		x			x		x
Screen 10	x			x	x		x			x		x	x			x
Screen 11	x			x	x			x	x			x	x			x
Screen 12	x			x	x			x		x	x			x		x
Screen 13	x			x		x	x		x		x			x	x	
Screen 14	x			x		x	x			x		x	x		x	
Screen 15	x			x		x		x	x			x	x		x	
Screen 16	x			x		x		x		x	x			x	x	
Screen 17		x	x		x		x		x		x			x		x
Screen 18		x	x		x		x			x		x	x			x
Screen 19		x	x		x			x	x			x	x			x
Screen 20		x	x		x			x		x	x			x		x
Screen 21		x	x			x	x		x		x			x	x	
Screen 22		x	x			x	x			x		x	x		x	
Screen 23		x	x			x		x	x			x	x		x	
Screen 24		x	x			x		x		x	x			x	x	
Screen 25		x		x	x		x		x		x			x		x
Screen 26		x		x	x		x			x		x	x			x
Screen 27		x		x	x			x	x			x	x			x
Screen 28		x		x	x			x		x	x			x		x
Screen 29		x		x		x	x		x		x			x	x	
Screen 30		x		x		x	x			x		x	x		x	
Screen 31		x		x		x		x	x			x	x		x	
Screen 32		x		x		x		x		x	x			x	x	

FA: Functional Areas:
Y = defined on model screen
N = no functional areas defined

DC: Directive Cues:
Y = present on model screen
N = no directive cues present

PI: Paragraph Indication:
Sp = line space indicates paragraphs
In = paragraphs indicated by indentation

HD: Headings:
Y = headings present on screen
N = no headings in text

LS: Line Spacing:
Ss = single spacing in text
Ds = double spacing between text lines

IL: Illustrations:
Y = illustration present on model
N = no illustration on model

SB: Status Bar:
Y = status bar present on model
N = no status bar on model

TC: Text Columns:
1 = single column of text on model
2 = two columns of text on model

FIGURE 3 □ Model Screens Used in Paired-Comparison Task

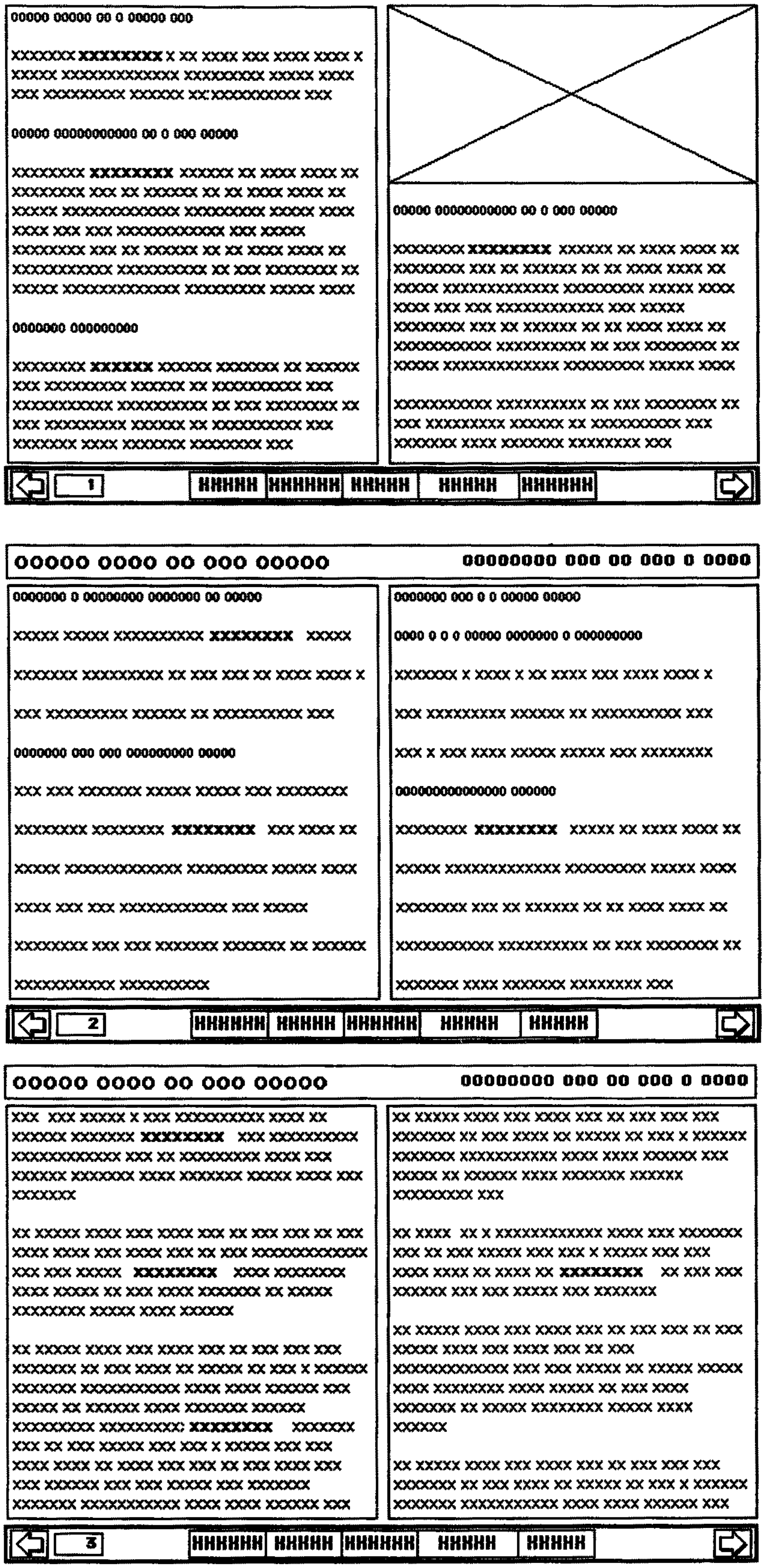


FIGURE 3 □ continued

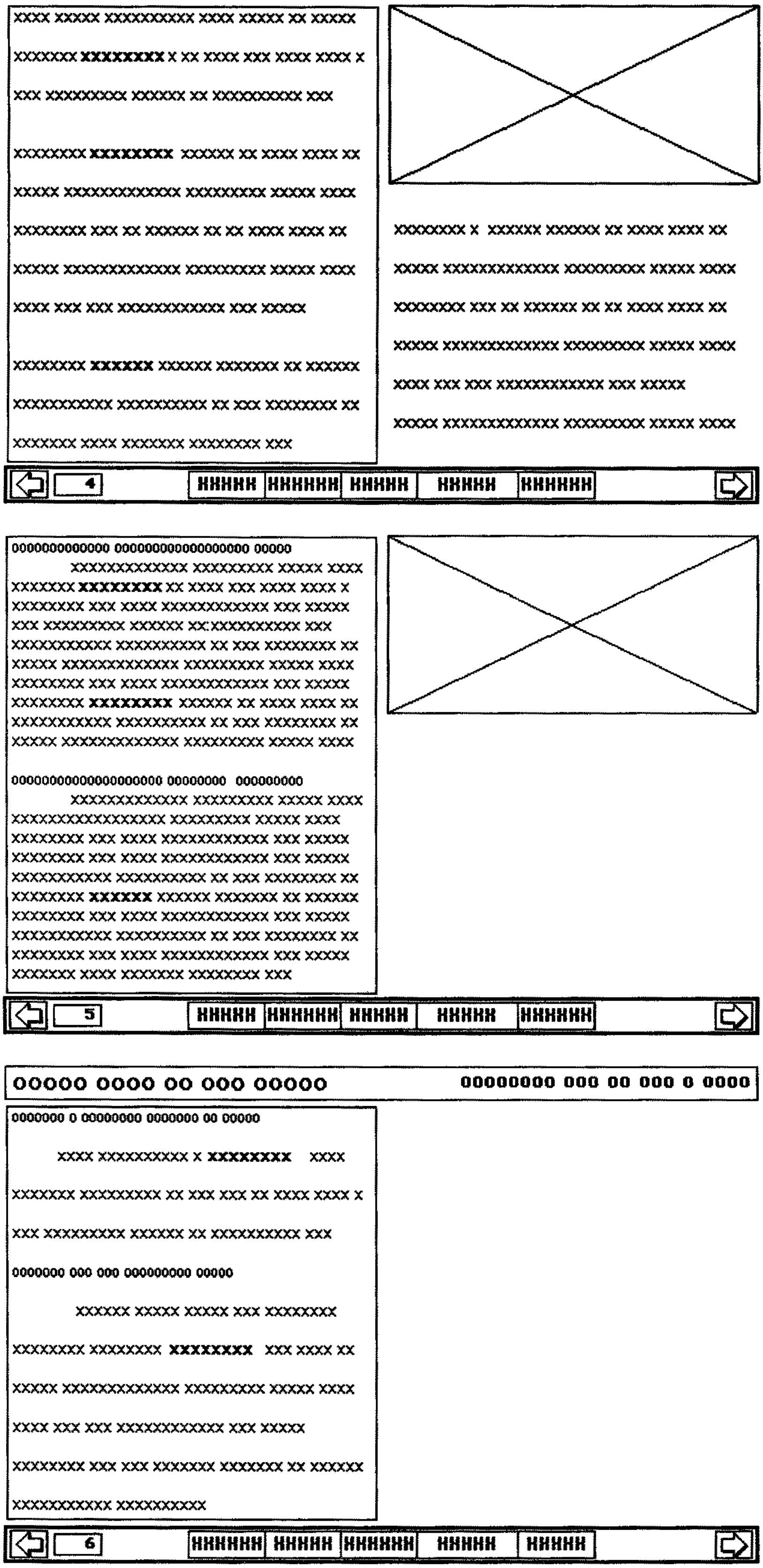


FIGURE 3 □ continued

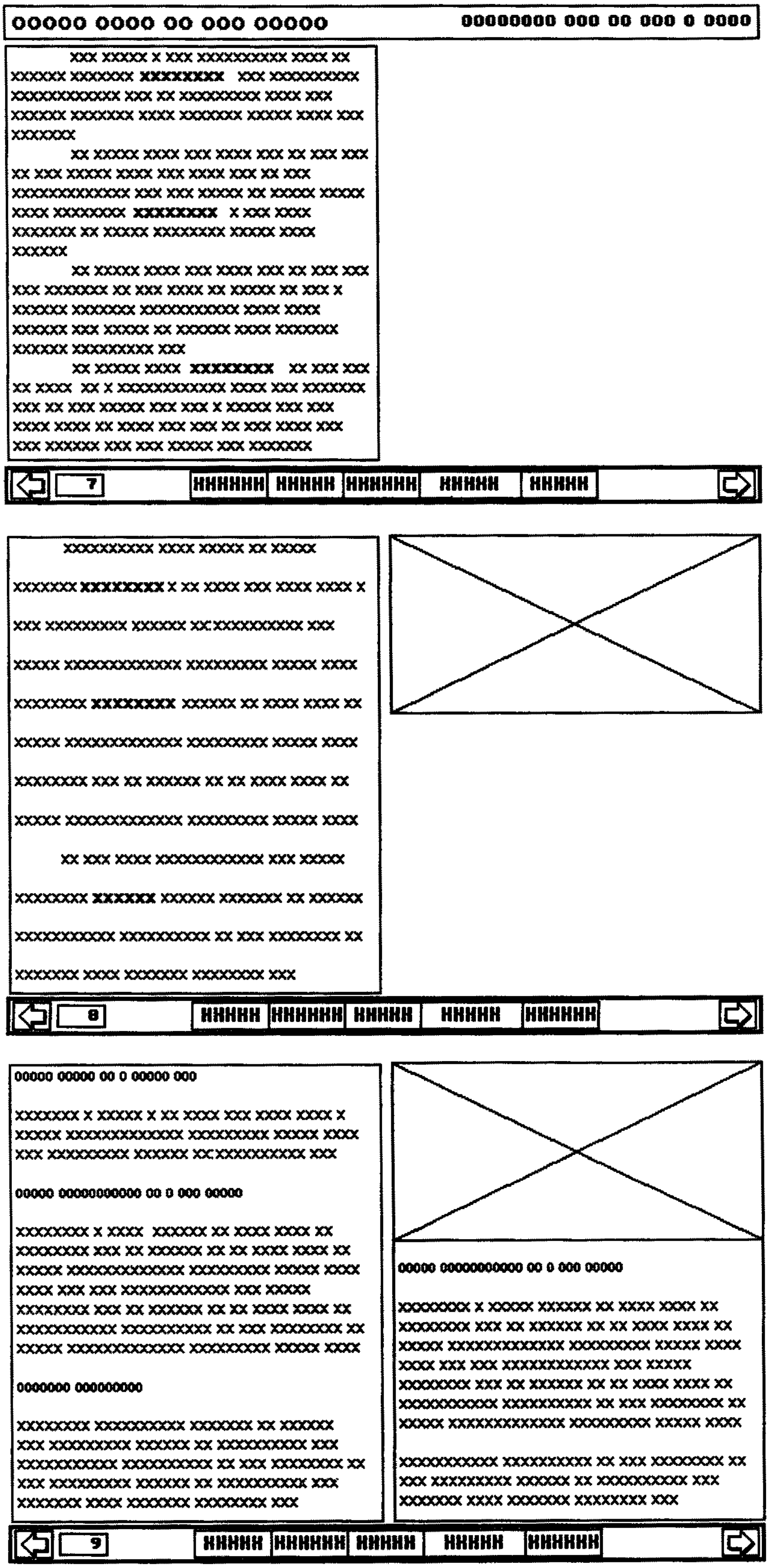


FIGURE 3 □ continued

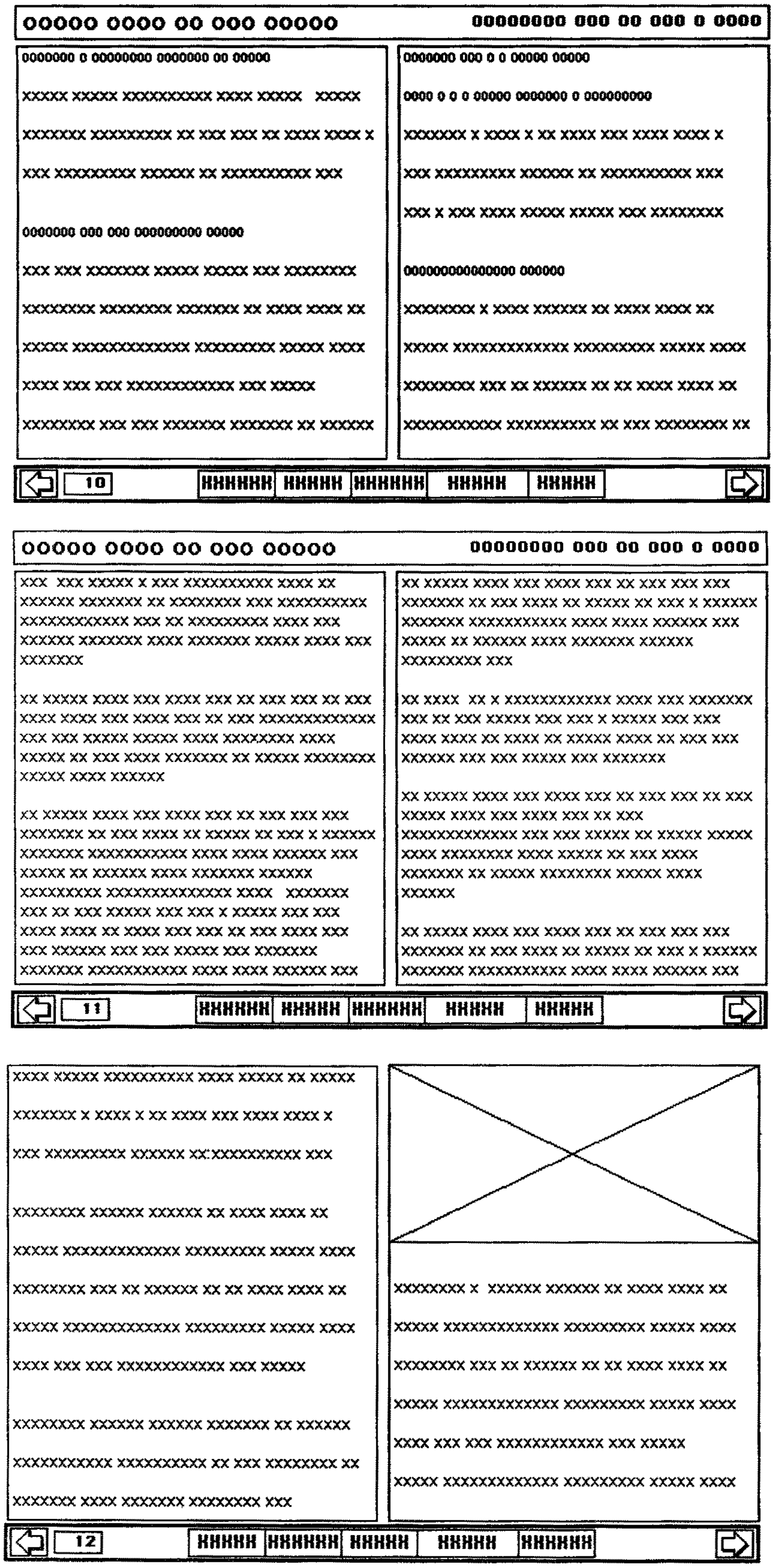


FIGURE 3 ☐ continued

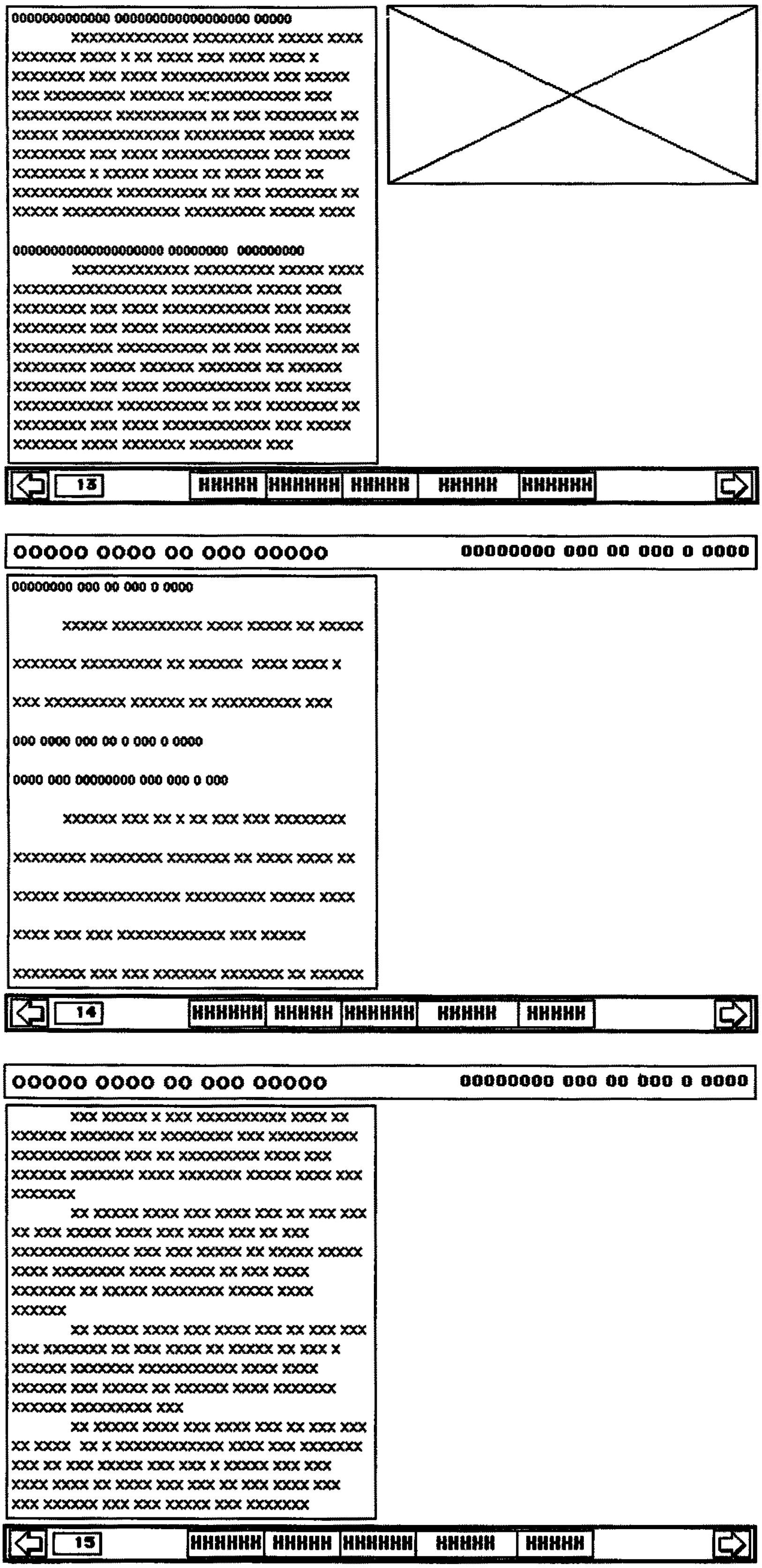
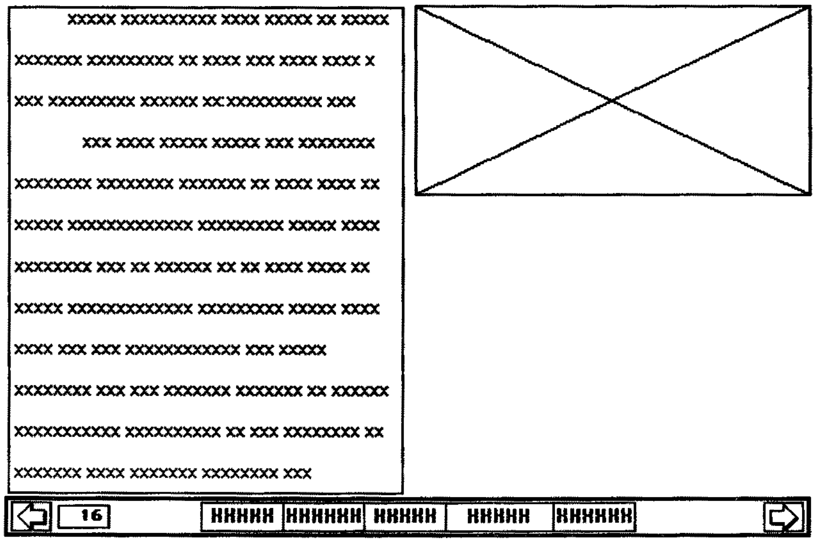
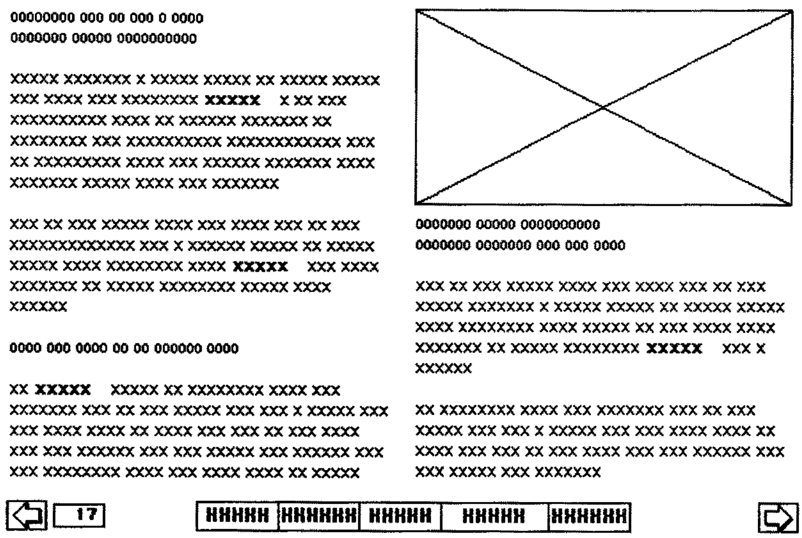


FIGURE 3 ☐ continued





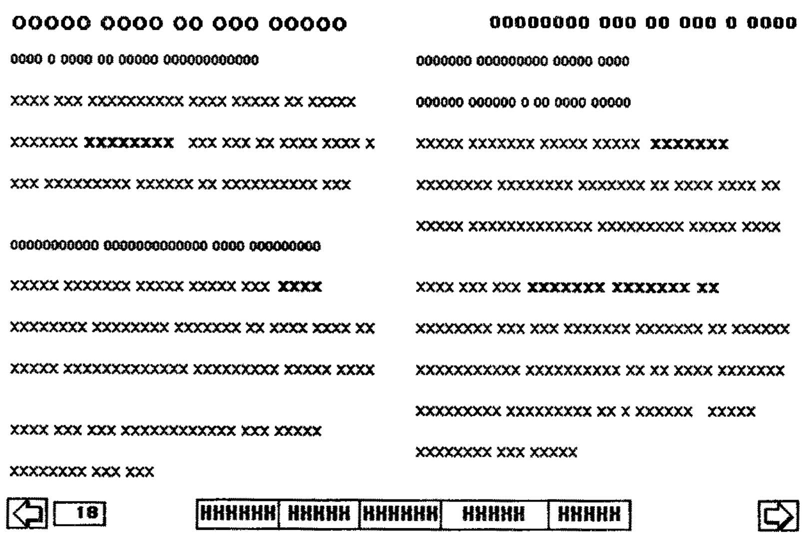


FIGURE 3 □ continued

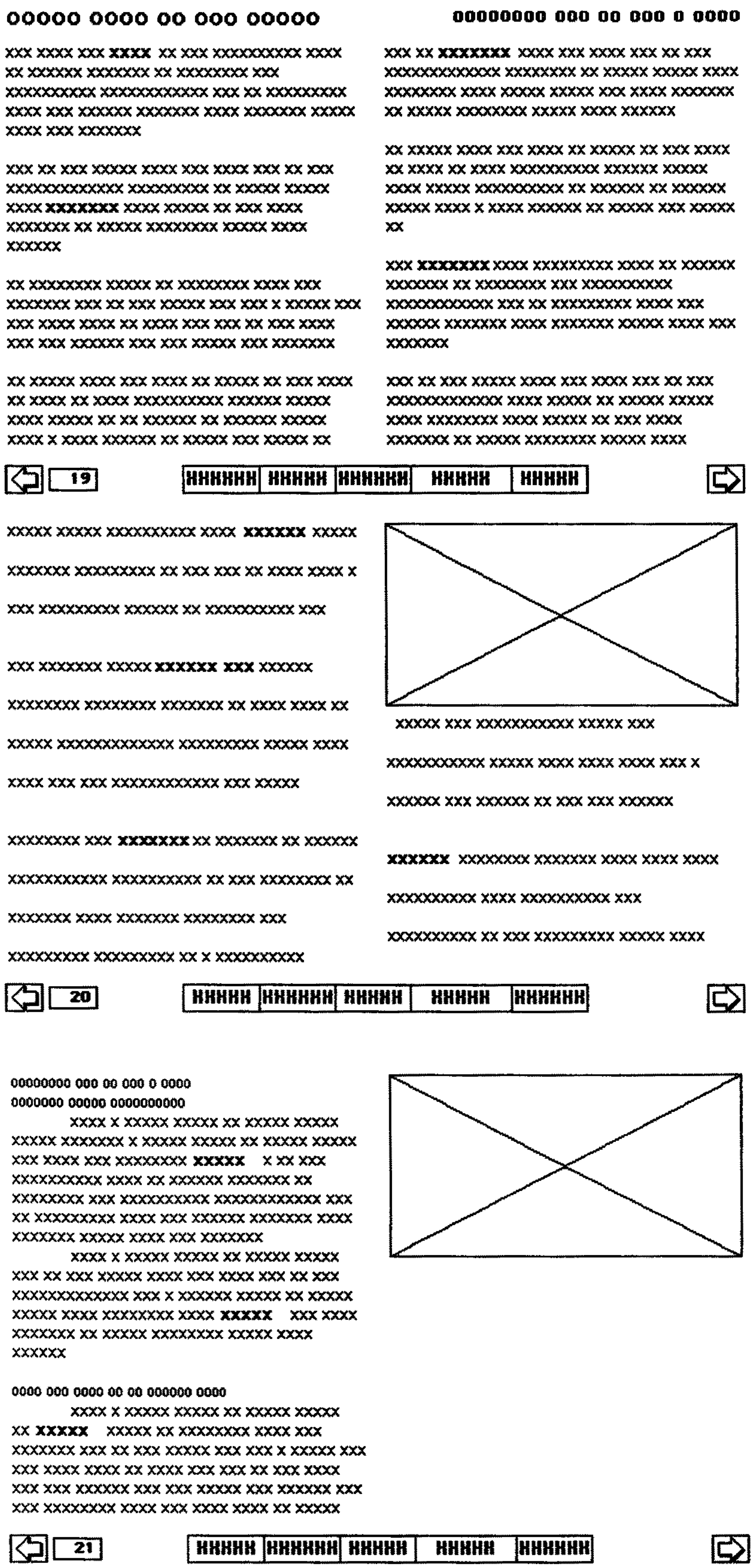


FIGURE 3 □ continued

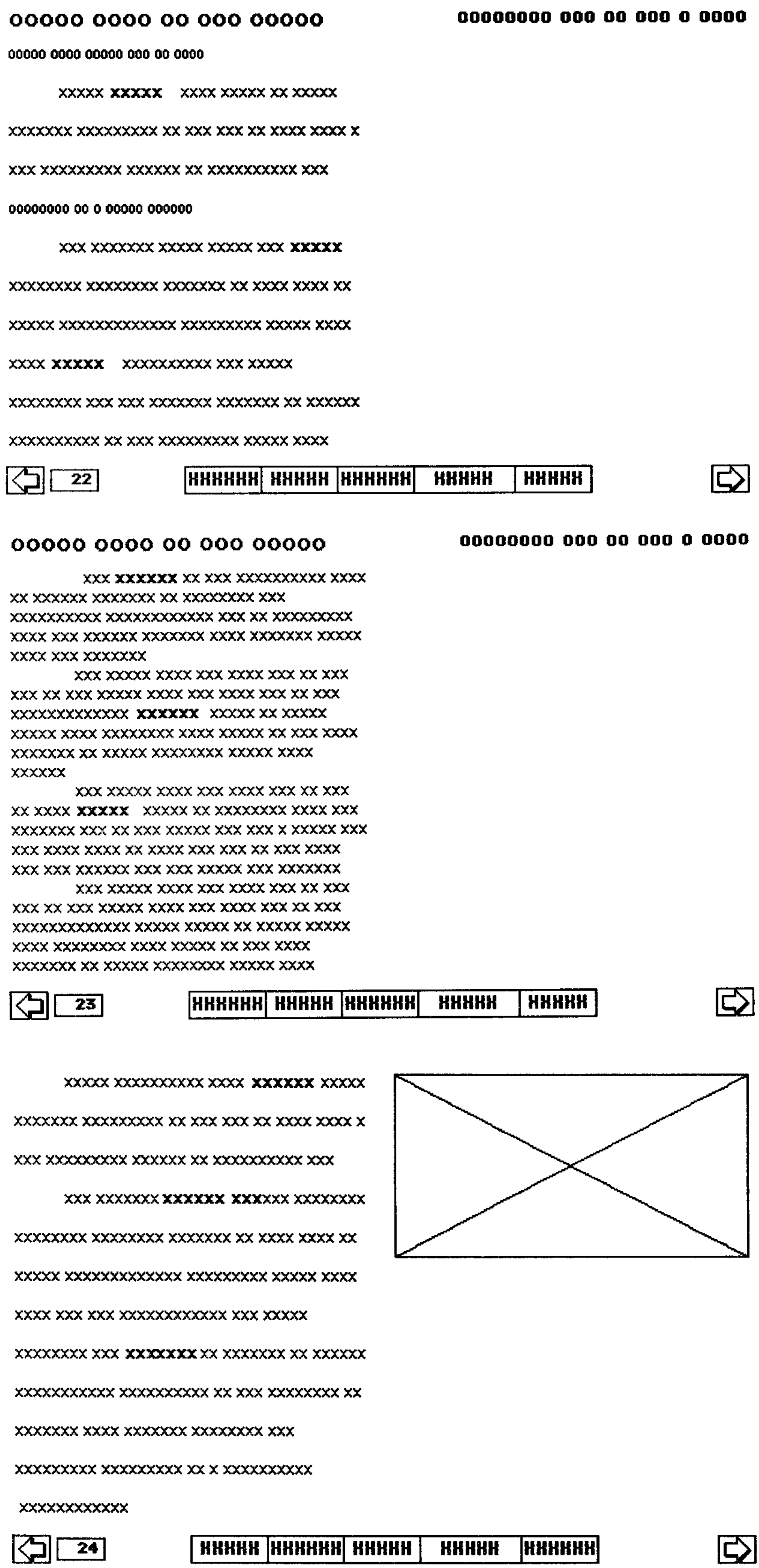


FIGURE 3 □ continued

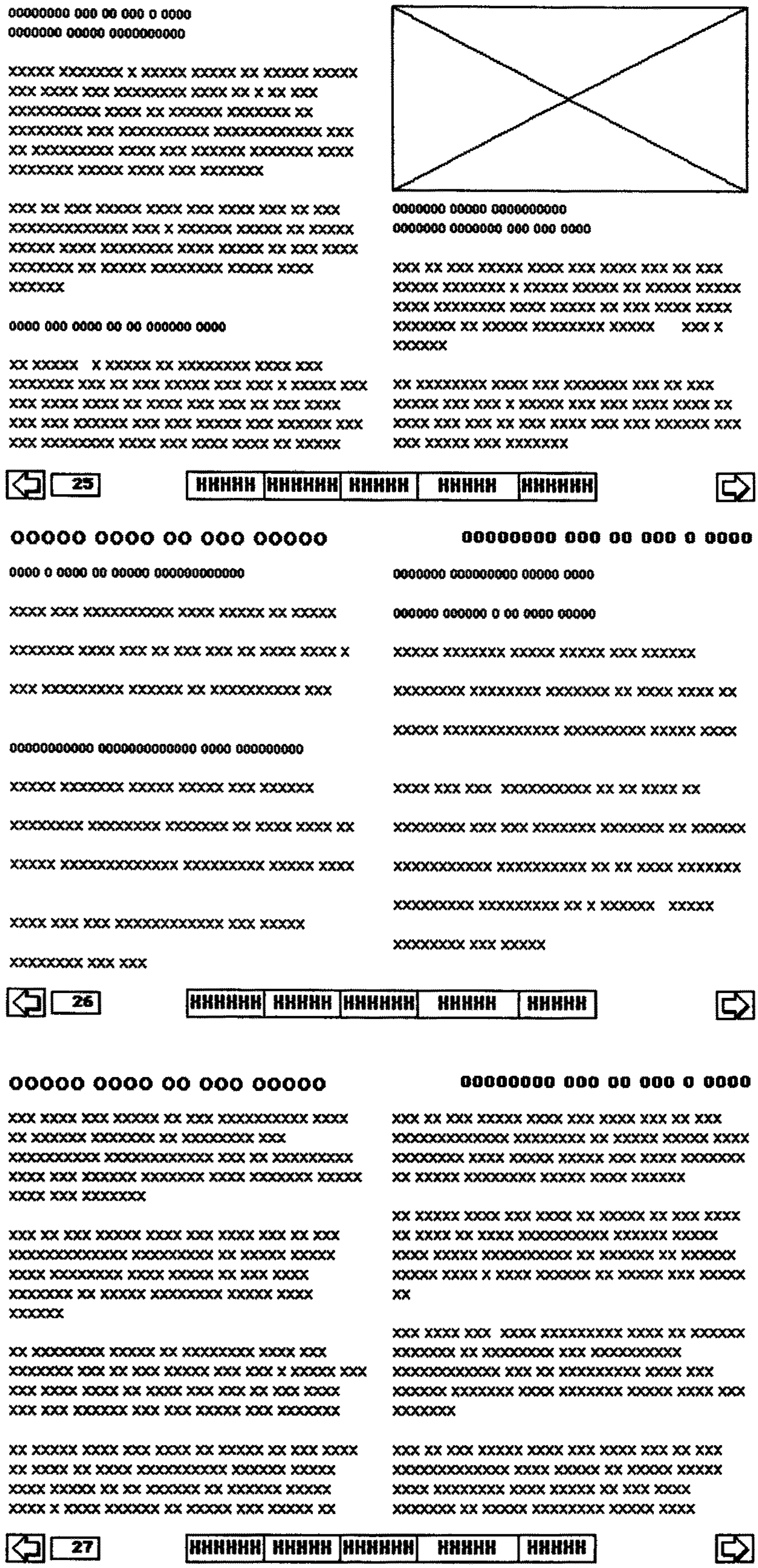
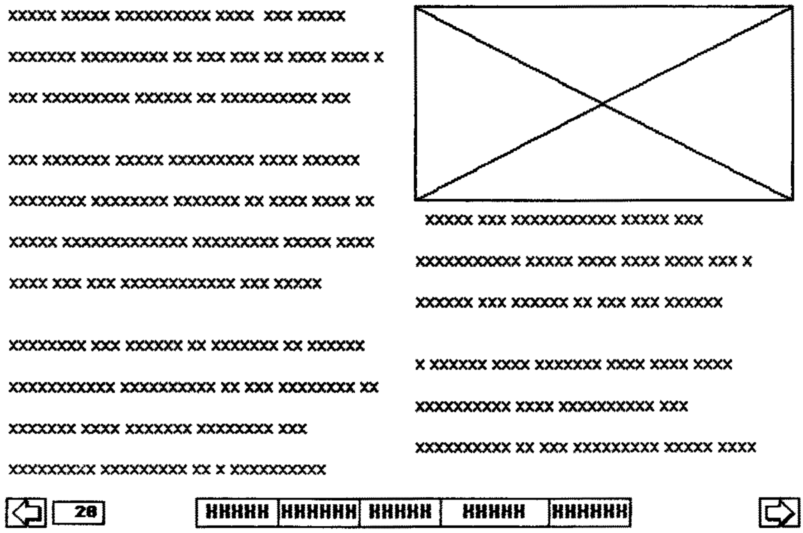
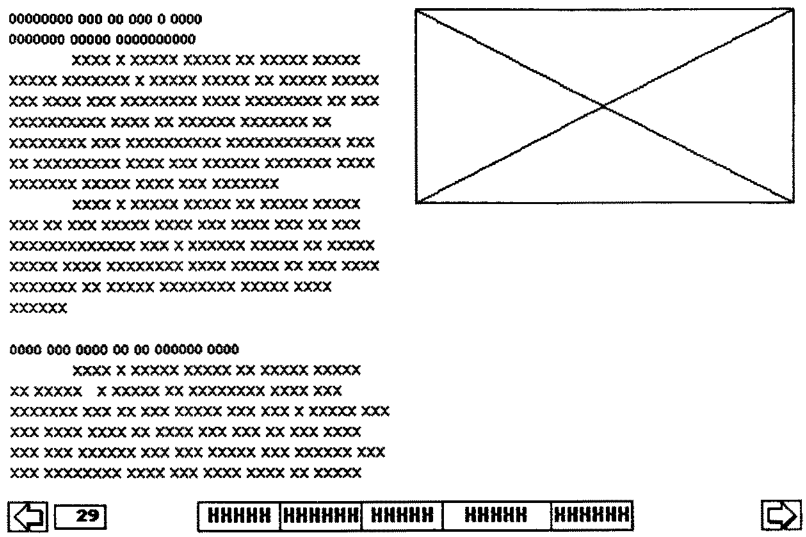


FIGURE 3 □ continued





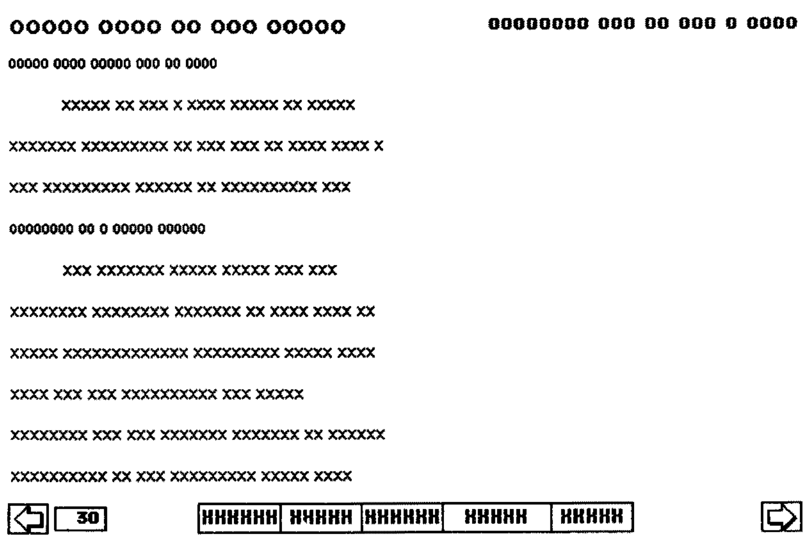
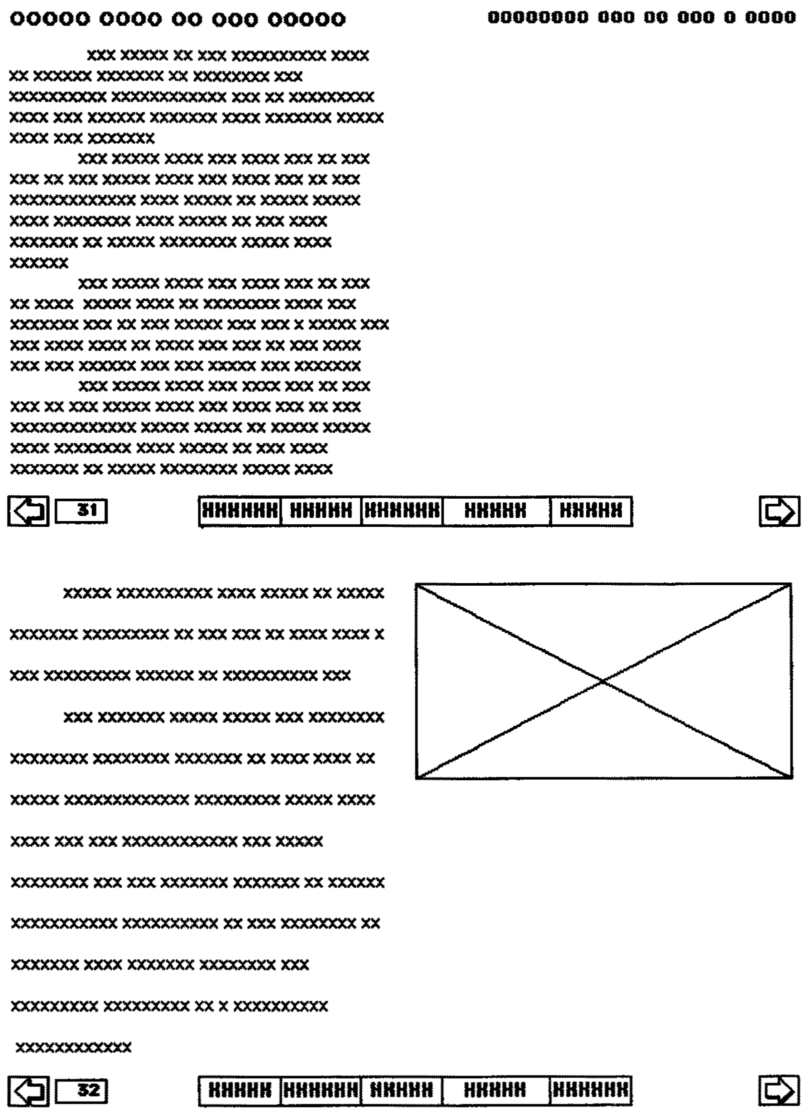


FIGURE 3 □ continued



Definition of Functional Areas (FA). Screen samples 1 through 16 exhibit this graphic technique. Rules separate the different functional areas of the screen; that is, the columns of text, the control panel at the bottom of the screen, and the status bar at the top of the screen (when present) are all enclosed in boxes. The two values were *present* (screens 1–16) and *not present* (screens 17–32).

Directive Cues (DC). Directive cues are elements that call attention to special words or areas of text. In this case, directive cues were "words" of text in bold type (see screens 1–8). Values for directive cues were *present* and *not present.*

Paragraph Indication (PI). Paragraphs were indicated in two ways: *increased space with no indentation* and *indentation with no space.* Indented paragraphs (screens 5–8) were indicated in the customary way of indenting the first line without any added space between paragraphs. Spaced paragraphs (screens 1–4) were indicated by adding space between paragraphs and not indenting the paragraphs.

Headings (HD). Headings were indicated by O's within the text (screen 1). Values for headings were *present* and *not present.*

Line Spacing (LS). Line spacing, or leading, methods were *single spacing* (screen 1) and *double spacing* (screen 2).

Illustrations (IL). Half of the screens exhibited a model illustration (screen 1), displayed as a large box with an X in it. Values for illustrations were *present* and *not present.*

Status Bar (SB). A status bar was placed along the top of the screen (screen 2). Participants were told that the status bar provided information about where in a lesson they were and about the lesson's main topics and subtopics—a necessity in most computer-based lessons or hypertext/hypermedia. Values for the status bar were *present* and *not present.*

Text Columns (TC). Text columns had two values: *one* (screen 1) and *two* (screen 6).

The $2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2$ design formed by the eight text element combinations presented 256 possible stimulus combinations. To reduce this to a workable number (32 screens), a fractional factorial design was used whereby each level appeared an equal number of times (Green, 1974). However, each factor did not appear an equal number of times under each other factor. For example, the illustration (IL) and status bar (SB) conditions are mirror images of each other, as are the text column (TC) and paragraph indication (PI) conditions. This means that when illustrations are present, the status bar is not, and that two columns of text are always single-spaced while a single column of text is always double-spaced.

Because of these interrelationships, no main effects for text element conditions were possible. Even 32 screen models leads to 496 [32 $\times$ 31)/2] discrete comparisons, so a predetermined group of screen pairs was shown to each subject. A cyclic design was used to determine the pairs used, with each subject viewing only a partial set of stimuli (Spence & Domoney, 1974). Each subject made only 62 comparisons, one-eighth of the total comparisons possible. For example, a subject was asked to judge the readability and studyability of screen 1 compared to screen 32, screen 2 compared to screen 16, and so on. The eight separate comparison routines were alternated with each subject to ensure that a balanced number of comparisons were made for each pair across all subjects.

In all, each comparison pair was viewed at least 11 times. (Kruskal & Wish, 1978, recommend that each comparison pair be viewed 10 times in a multidimensional scaling study.) Each comparison pair could then be labeled, for example, pair 1, pair 2, pair 3, and so on, up to pair 496. The first subject viewed pairs that composed Routine 1, or pairs 1 to 62. The second subject judged the pairs in Routine 2, pairs 63 to 124, and so on for each of the eight routines.

Adjective Rating Task and Materials

Recall that the MDS procedures separate the stimuli into similar and dissimilar groups but do not interpret the reasons for separating the stimuli. So, to aid in interpretation of the MDS results, the subjects "described" the screens using an adjective differential scale. The intent of this task was to determine whether any of the adjectives loaded onto the MDS screen coordinates through a multiple regression analysis. If one or more adjectives regressed significantly on one of the dimensions, those adjectives were said to describe that dimension and lend support to the visual interpretation. So, if the term "organized" regressed significantly on dimension 1, stimuli at one end of the dimension were expected to be more organized than those at the other end.

Adjectives on the rating form were selected before the study by asking a group of viewers to describe a selection of model screens. Their comments formed the basis for an initial group of 20 adjectives, which were reduced to the final total of 13 through discussion among researchers (see Figure 4). In this case, because of the large number of screens (32),

FIGURE 4 □ Adjective Rating Scale

messy	neat
dirty	clean
cluttered	organized
crowded	spacious
unattractive	attractive
unplanned	planned
unreadable	readable
hard to study	easy to study
boring	interesting
unstructured	structured
intimidating	inviting
static	dynamic
accidental	controlled

the paired-comparison task was completed first. Next, a sample of 10 screens that represented the extremes of the MDS dimensions were selected. Finally, a random sample of 40 subjects was brought back to rate each of the 10 screens on the adjective scale (one form for each screen).

Procedures

Study 1 used a paired-comparison test as the primary treatment. Each subject sat facing two Macintosh SE computers. All subjects filled out an information form and signed a release. They then listened to recorded instructions which told them to examine and compare each screen according to "how much [they] would like to read and study from the screen if it were an actual screen in a computer lesson." Participants practiced with a set of six comparisons to get used to selecting the left or right image and marking the form; then the actual task began. Two separate screen samples were presented simultaneously for seven seconds. After choosing the preferred screen (left or right), participants circled the letter L or R on the response form (Figure 2) and made a mark on the scale reflecting how much they preferred one screen over the other. The next pair of screens was then presented, and the procedure continued until the subject had compared a total of 62 screens.

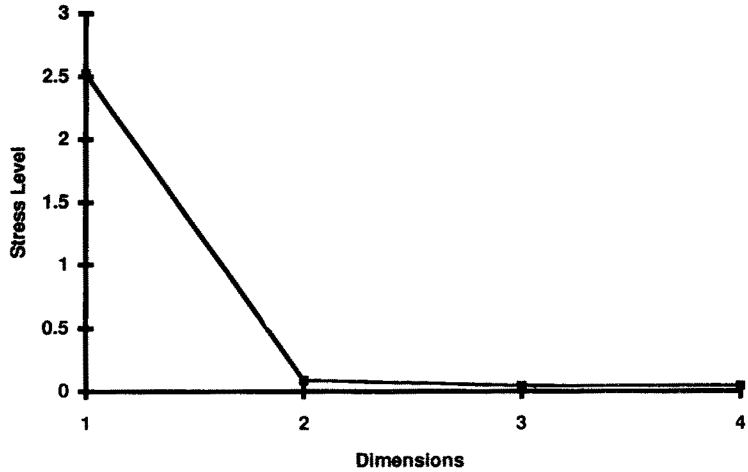
As a final activity, subjects rated screens on the 13-item adjective scale (Figure 4). A single screen was shown on a monitor and the subject examined the screen while marking the lines associated with the adjectives on the form to describe that screen. This process continued until the participants had evaluated a representative sample of 10 screens.

Results

Results from the multidimensional scaling analysis indicated that participants evaluated the stimuli along a common set of two dimensions. The MDS interpretation was limited to two dimensions because an extremely low level of stress was arrived at after 44 iterations (.088). *Stress* is a measure of the goodness-of-fit of the calculated solution; the lower the stress, the more likely the solution represents reality. (Technically, stress is the square root of a normalized residual sum of squares.) Figure 5 presents the stress reduction that occurred when the participant utilities were run through one-, two-, three-, and four-dimensional solutions.

The two-dimensional solution was accepted for interpretation because there was a large change between the one-dimensional (.252) and two-dimensional solution (.088), but little change between the two-dimensional solution (.088) and the three-dimensional (.046) and four-dimensional (.044) solutions. The result favoring the two-dimensional solution means that there seem to be two constructs or criteria on which viewers judge the readability and studyability of model screens. Although the stress level (.088) is very low and indicates that the empirical solution is valid,

FIGURE 5 □ MDS Stress Values for Model Screens along Four Dimensions



its stability is not perfect (perfect stress = .000), implying that there are unexplained sources of variance within the data. The effect of this instability or inaccuracy is to make the interpretation of the dimensions more difficult because some screens may be "forced" into a group in the attempt to create a model with the best fit. Items forced into groups may not exhibit the characteristics expected as strongly as other items in the group.

Figure 6 presents the MDS results for the model screens in one two-dimensional figure. The actual two-dimensional coordinates for the MDS dimensions are found in Table 4. MDS techniques have no built-in procedures for labeling the dimensions. The first step in interpreting MDS results is to examine the stimuli locations on the maps and axis of each dimension in terms of the objective, visible characteristics of the stimuli. The proximity of the stimuli along each axis represents the extent to which one screen is perceived as similar or different from the others. Within each dimension, a few stimuli that are perceived as unique usually stand out separate from the others, such as screens 1, 3, 4, and 5 in Figure 6.

FIGURE 6 □ MDS Configuration for Model Screens along Two Dimensions

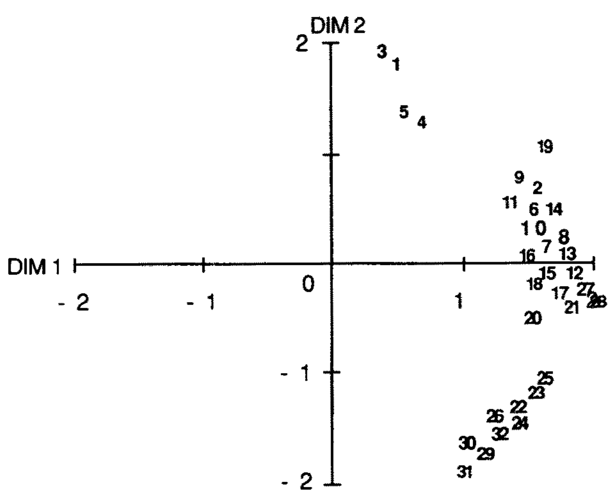


TABLE 4 □ Two-Dimensional Coordinates for Real Screens

Screen	*Dimension 1*	*Dimension 2*	*Screen*	*Dimension 1*	*Dimension 2*
SCR01	0.40	1.73	SCR17	1.81	−0.19
SCR02	1.57	0.52	SCR18	1.76	−0.06
SCR03	0.34	1.80	SCR19	1.65	0.52
SCR04	0.76	1.59	SCR20	1.73	−0.61
SCR05	0.66	1.63	SCR21	1.75	−0.27
SCR06	1.51	0.44	SCR22	1.46	−1.36
SCR07	1.69	0.09	SCR23	1.46	−1.44
SCR08	1.71	0.16	SCR24	1.35	−1.52
SCR09	1.44	0.81	SCR25	1.56	−1.33
SCR10	1.68	0.37	SCR26	1.46	−1.46
SCR11	1.60	0.50	SCR27	1.86	−0.30
SCR12	1.80	−0.10	SCR28	1.80	−0.34
SCR13	1.77	0.01	SCR29	1.23	−1.79
SCR14	1.68	0.29	SCR30	1.23	−1.79
SCR15	1.79	−0.06	SCR31	1.12	−1.97
SCR16	1.76	0.07	SCR32	1.51	−1.39

Another step in interpretation is the ranking of utility means of the stimuli to indicate those that are the most studyable and readable (see Table 5). The stimuli rankings are an average of each participant's utilities (normalized rankings) for each screen, and therefore indicate which screens the subjects ranked highest on the factors of studyability and readability. These rankings help determine whether one end of a dimension represents those samples that the viewers judged more readable and studyable or those that the viewers judged less so. Note that the purpose of multidimensional scaling is not to identify the screens that subjects liked most, but to sort the screens by judgments made by the viewers. This procedure only gives an indication of which end of a dimension may be more preferred than the other. Because the MDS dimensions reflect a statistical effort to group stimuli by the most salient characteristic, not necessarily by how much they are preferred, it is likely that disliked stimuli are scattered among the preferred stimuli.

Finally, to help interpret the MDS dimensions, adjectives with which the participants rated the screens were regressed on the MDS coordinates. The first step in analyzing the ad-

TABLE 5 □ Ranking of Model Screens by Mean

Rank	*Screen*	*Mean*	*Rank*	*Screen*	*Mean*
01.	SCR25	55.8	17.	SCR14	42.2
02.	SCR17	53.7	18.	SCR13	42.0
03.	SCR09	52.8	19.	SCR03	41.8
04.	SCR01	51.9	20.	SCR02	41.4
05.	SCR04	49.5	21.	SCR06	41.0
06.	SCR20	48.1	22.	SCR24	40.8
07.	SCR21	47.7	23.	SCR26	40.7
08.	SCR08	47.3	24.	SCR27	40.4
09.	SCR11	46.6	25.	SCR23	40.3
10.	SCR12	46.4	26.	SCR16	40.2
11.	SCR29	46.2	27.	SCR18	40.1
12.	SCR10	46.1	28.	SCR30	40.1
13.	SCR28	43.8	29.	SCR07	38.8
14.	SCR32	42.8	30.	SCR19	38.4
15.	SCR05	42.7	31.	SCR22	36.9
16.	SCR15	42.3	32.	SCR31	35.5

Note: The higher the mean, the more the screen was preferred.

jective scale data was the creation of a correlation matrix (Table 6). The correlation matrix shows a high degree of intercorrelation among the adjectives. The intercorrelations suggest that many of the adjectives were perceived as similar by the subjects and were poor discriminators. A multiple regression based on this matrix would have little meaning because the variance would be split among highly similar items. Therefore, using Statview 512+ (Brainpower, Inc., 1986), a principal components factor analysis using an oblique rotation was conducted to create a more distinct set of descriptors.

An oblique rotation was used because of the intercorrelation among adjective ratings. The oblique rotation "is more desirable because it is theoretically and empirically more realistic" (Hair et al., 1979, p. 225) in developing constructs to explain the adjective ratings. Note that oblique rotations report two sets of factor loadings: pattern loadings and structure loadings. Pattern loadings are the "measures of the unique contribution each factor makes to the variance of the variables" (Rummel, 1970, p. 397). The structure loadings are "the product moment correlations of the variables with the oblique factors" (p. 399) and represent both the loading of the variable and the interactions of other factors with that factor. The pattern loadings are best used for interpreting the meaning of the factor because they indicate the unique contribution of a variable to a factor and define the simple structure configuration (Rummel, 1970). The squared structure loadings, on the other hand, give a measure "of the variance of each variable jointly accounted for by a factor and the interaction effects of that factor with others" (p. 401). However, because the structure loadings represent the loading of a variable on the factor and the interaction of every other factor with that variable, the interpretative value of the structure loadings is minimal.

Table 7 shows the pattern and structure loadings for the present study. Using the pattern loadings, the adjectives that loaded on Factor 1 were: "neat," "clean," "organized," "spacious," "attractive," "readable," "studyable," and "inviting." Factor 1 is composed of adjectives that lead to a neat and organized screen and is therefore called "organized." The adjectives that loaded on Factor 2 were: "planned," "interesting," "structured," "dynamic," and "controlled." Factor 2 is composed of adjectives that refer to a more microarrangement and composition of the screens and is referred to as "structured."

Next, the factor scores resulting from the analysis were regressed onto each dimension to see if they could help explain the dimension and lend support to the visual analysis. The results of the multiple regression showed that the factor labeled earlier as organization loaded significantly on the first dimension ($t = 2.582$, $p = .01$). Neither factor loaded on the second dimension, though the organization factor approached significance ($p = .053$).

TABLE 6 ☐ Model Screen Adjective Correlation Matrix

Adjectives	*1.*	*2.*	*3.*	*4.*	*5.*	*6.*	*7.*	*8.*	*9.*	*10.*	*11.*	*12.*	*13.*
1. Neat	1.00	.90*	.76*	.53	.63*	.62*	.65*	.60*	.55	.56	.55	.50	.53
2. Clean		1.00	.77*	.61*	.67*	.62*	.67*	.61*	.58*	.59*	.62*	.54	.55
3. Organized			1.00	.76*	.74*	.66*	.77*	.76*	.65*	.54	.71*	.56	.56
4. Spacious				1.00	.68*	.55	.67*	.67*	.65*	.45	.71*	.59*	.49
5. Attractive					1.00	.70*	.76*	.79*	.81*	.61*	.74*	.66*	.63*
6. Planned						1.00	.73*	.67*	.66*	.76*	.64*	.58*	.77*
7. Readable							1.00	.83*	.69*	.57*	.74*	.61*	.63*
8. Studyable								1.00	.76*	.55	.79*	.71*	.59*
9. Interesting									1.00	.62*	.78*	.74*	.61*
10. Structured										1.00	.58*	.56	.78*
11. Inviting											1.00	.77*	.63*
12. Dynamic												1.00	.61*
13. Controlled													1.00

*$p < .05$.

TABLE 7 ☐ Oblique Factor Analysis of Adjectives for Model Screens

	PATTERN LOADINGS		STRUCTURE LOADINGS	
Adjectives	*Factor 1*	*Factor 2*	*Factor 1*	*Factor 2*
1. Neat	.86*	−.02	.62	−.02
2. Clean	.87*	−.00	.63	−.00
3. Organized	.92*	.00	.67	.00
4. Spacious	.81*	.03	.59	.02
5. Attractive	.57*	.39	.41	.29
6. Planned	.15	.77*	.11	.56
7. Readable	.64*	.31	.47	.22
8. Studyable	.61*	.34	.45	.24
9. Interesting	.39	.54*	.28	.39
10. Structured	−.11	.95*	−.08	.70
11. Inviting	.50*	.45	.36	.32
12. Dynamic	.28	.58*	.20	.42
13. Controlled	−.11	.98*	−.08	.71

*Adjective loads on this factor.

Note: The factor procedure used was principal components analysis using an oblique solution primary pattern matrix transformation method.

With this information and the rankings and spatial map of MDS coordinates, the dimensions could then be interpreted.

Dimension 1: Organization

The first dimension was initially interpreted by visually comparing screens that represented extremes along the horizontal dimension: screens 1, 3, 4, and 5 against screens 7, 8, 12, 13, 15, 17, 27, and 28. The remaining screens were not selected because they seemed to be perceived by subjects as basically similar. Screens 1, 3, 4, and 5 were viewed slightly more positively (mean ratings = 46.5) than the screens in the second group (mean ratings = 44.3). Analysis of the characteristics of the first group of screens (1, 3, 4, and 5) showed that all four had defined functional areas and directive cues. Three of the four used spacing to indicate paragraphs and had single-spaced lines and double text columns. The overall effect of these characteristics is to present a balanced, organized appearance. The second group of screens (7, 8, 12, 13, 15, 17, 27, and 28) uses a single text column in half the samples and fewer graphic devices, presenting a lighter, less cluttered appearance and a lack of aesthetic balance.

Only three of the eight screens use boxes to define functional areas. Half of the screens use spacing and half use indentation for paragraph indication. Half of the screens use directive cues and only two of the screens (13 and 17) use headings. As reported in the previous section, the first adjective factor, organization, also loaded heavily on the first dimension. Viewers appeared to notice the gross arrangement of the screen, focusing on the lines that define functional areas and the spaced paragraphs, elements that create screens that can be described as organized, readable, and studyable.

Dimension 2: Visual Interest

The second dimension (vertical axis in Figure 6) is more obvious in differences among visual characteristics than the first. The second dimension was interpreted visually by comparing screens 1, 3, 4, and 5 against screens 24, 26, 29, 30, 31, and 32. Note that the former group is the same set of screens as in the first dimension; it is not unusual to have stimuli share features of both dimensions. The dimensions in multidimensional scaling are not orthogonal but share features to a greater or lesser degree—in this case, to a greater degree. Again, screens 1, 3, 4, and 5 were rated higher (mean ratings = 46.5) than screens 24, 26, 29, 30, 31, and 32 (mean ratings = 41.0). The screens in the latter group are much more open, plainer, and less interesting than those in the former group. Four of the six are double spaced and have no illustration. Five of the six use only one text column. None of the screens has functional areas defined.

As reported previously, neither of the adjective factors regressed significantly on the second dimension, although the first adjective, organization, approached significance ($p = .0528$; see Table 9). This is not unusual, since the same four screens represent the positive end of the dimension. Hence, screens 1, 3, 4, and 5 are judged as more studyable and readable than the others. To the viewers, these screens had the best combination of text elements, providing visually attractive, interesting, and intriguing screens.

Field Articulation, Conceptual Style, and Gender

Field Articulation. The field articulation results were split into four categories: high field dependence (those that fell into first quartile of the total possible score), low field dependence (second quartile), low field independence (third quartile), and high field independence (upper quartile). The MANOVA compared the categories against the average utility score for each screen and showed no significant differences among the categories (Pillais $F = 1.212$, $p = .194$) across any of the screens.

Conceptual Style. The conceptual styles test produced two scores: a score representing analytical thinking and a score representing relational thinking. Both scores were analyzed. First, the analytical scores were split into three categories: low analytical (lower 33% of total possible score), moderate analytical (mid 33%), and high analytical (upper third). The categories were then compared to the utility screen score for each screen. The MANOVA showed a significant difference ($F = 1.475$, $p = .035$) overall. However, when the conservative Scheffé test was used to compare the three analytical score means within each screen, no specific differences were found, nor was any trend discernible indicating that the MANOVA may have picked up a difference in a single screen. A single difference, or even a few, does not give enough information to draw any conclusions about subjects' analytical styles and their perceptions of screens.

After examining the analytical scores, the relational scores were examined in the same way: low, moderate, and high relational scores. Again, the MANOVA showed an overall significant difference (Pillais $F = .962$, $p = .008$). The Scheffé post hoc comparison found one significant difference ($p < .05$) in screen 13 (low relational = .055, high relational = .145). However, a difference within one screen out of 32 screens is probably an anomaly and does not provide the basis on which to draw any conclusions related to how relational thinkers view the screens.

Gender. Finally the screen utility scores were again used as the dependent variable in a MANOVA comparing male and female judgments for the 32 screens. No significant difference (Pillais $F = .350$, $p = .562$) was found between gender within any of the screens.

STUDY 2

Method

In most respects, Study 2 used the same methodology as Study 1. Differences are noted in the following discussion.

Participants

Participants in Study 2 were 33 graduate and undergraduate students from the University of Colorado at Denver. (See Table 2 for a summary of participant characteristics.) Subjects' ages ranged from the early 20s to over 60. There were 23 males, 7 females, and 3 subjects who did not report gender. Subjects were members of a variety of instructional technology classes and received credit for participating in the study. None of the subjects participated in Study 1 and none was familiar with screen design concepts.

Primary Task and Materials

While Study 1 was highly controlled in that it used eight text elements to create a series of model screens, Study 2 was less controlled so as to expand the definitions and generalizability of the constructs identified in Study 1. The models employed in Study 1 created screens that were free from content, therefore enabling subjects to focus on the overall design of the screen (Grabinger, 1984, 1987; Grabinger & Amedeo, 1988). However, though free from content, model screens are not very generalizable to the real world. Also, models built from a limited number of text elements show relatively little variation among most of the screens.

In contrast, Study 2 examined viewer judgments of paired comparisons of real screens copied from existing programs. An additional step not used in Study 1 began with the selection of 40 screens from a variety of Macintosh instructional computer tutorials. The author sampled screens that represented a wide variety of design characteristics, from simple to complex and from plain to sophisticated. Screens with many graphic features and a sophisticated design appearance were chosen, as well as ones with few text elements and little sophistication. Next, the 40 screens were copied onto paper and shown to 20 graduate students (these students did not participate in the paired-comparison task). The students were asked to judge the studyability and readability of each screen and place a mark on a 100-millimeter line to indicate their judgments. The ratings of each of the 40 screens were averaged and the 10 highest rated and 10 lowest rated screens were chosen for the paired-comparison task. The intent was to identify a set of highly different screens to help emphasize underlying judgments. The 20 screens are shown in Figure 7. Aside from this added step, Study 2 proceeded in the same manner as Study 1.

Procedures

Task completion and administration proceeded in the same way as Study 1. Even though there was a smaller number of screens, as in Study 1 an incomplete matrix design had to be used to prevent subject fatigue. However, due to the smaller number of screens, the comparison routine differed. Twenty screen displays have 190 discrete comparisons [(20 × 19)/2)], so each participant made 95 comparisons. Eight separate comparison routines, each with 95 pairs, were created to vary the order of the comparison pairs across the participants. As in Study 1, a cyclic routine was used (Spence & Domoney, 1974). The eight routines were alternated to ensure that a balanced number of comparisons were made for each pair of screens. In all, each unique pair of stimuli was viewed at least 15 times.

FIGURE 7 ☐ Real Screens Used in Paired-Comparison Task

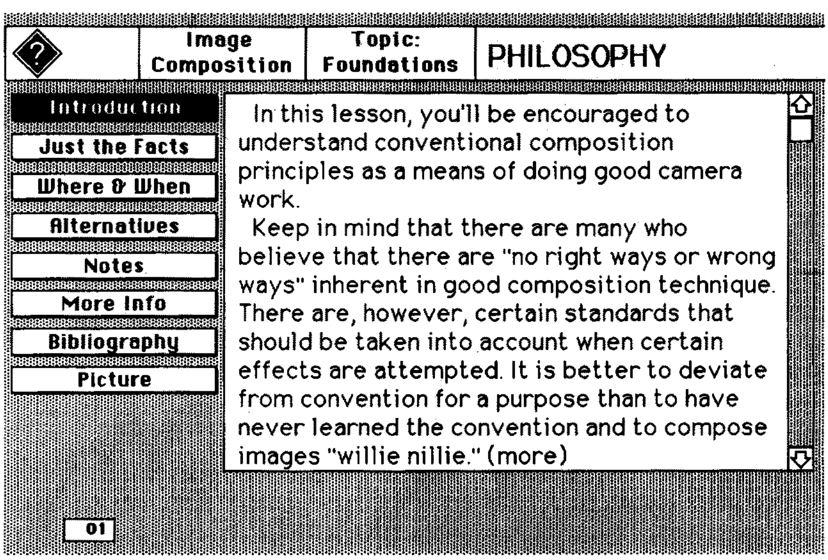
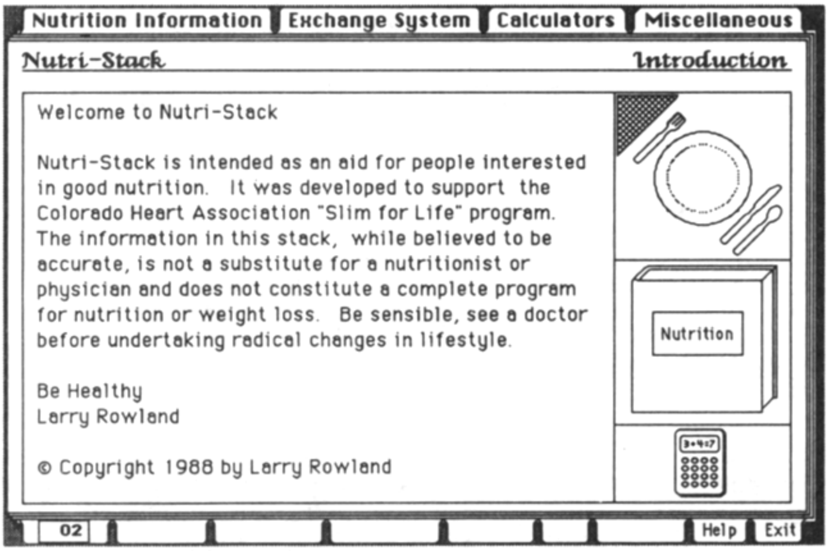
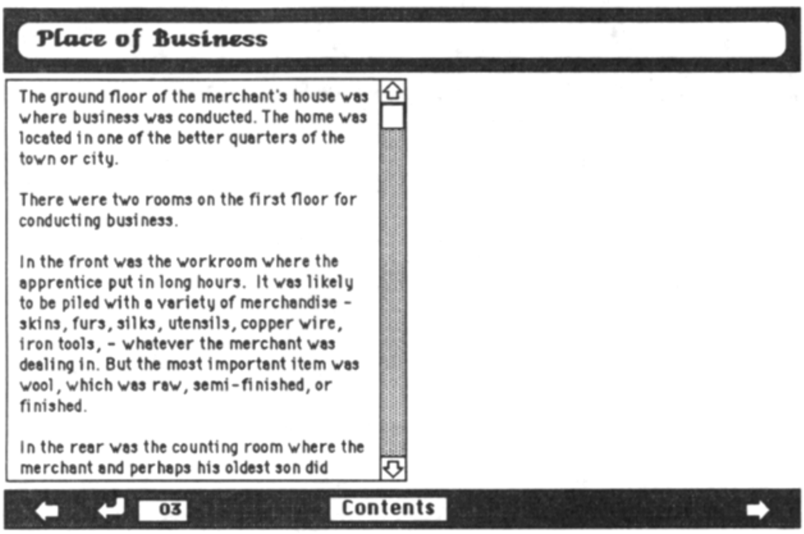


FIGURE 7 □ continued





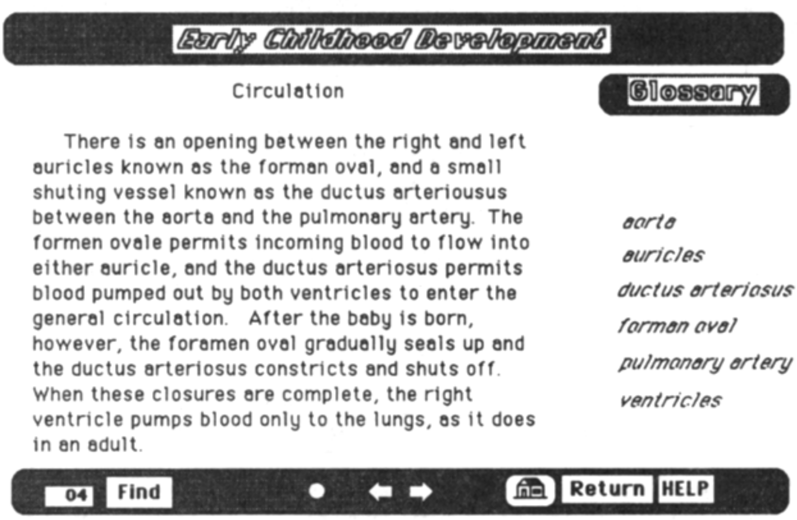
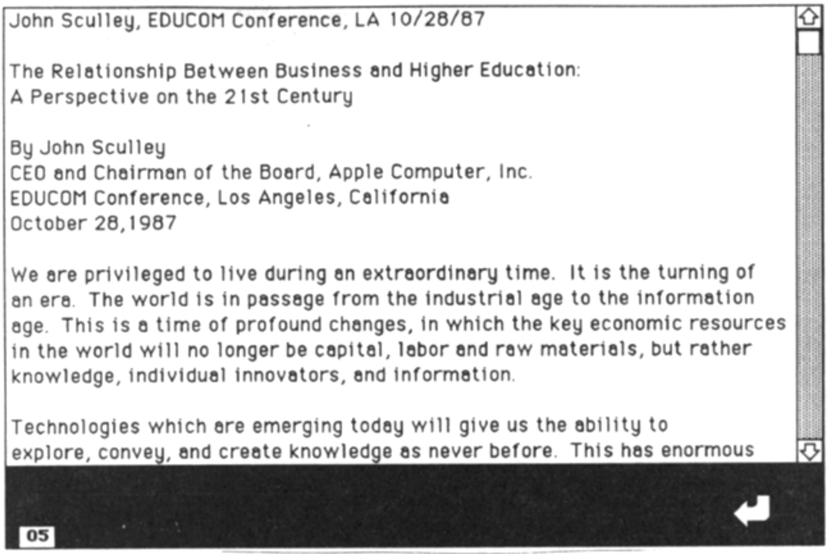
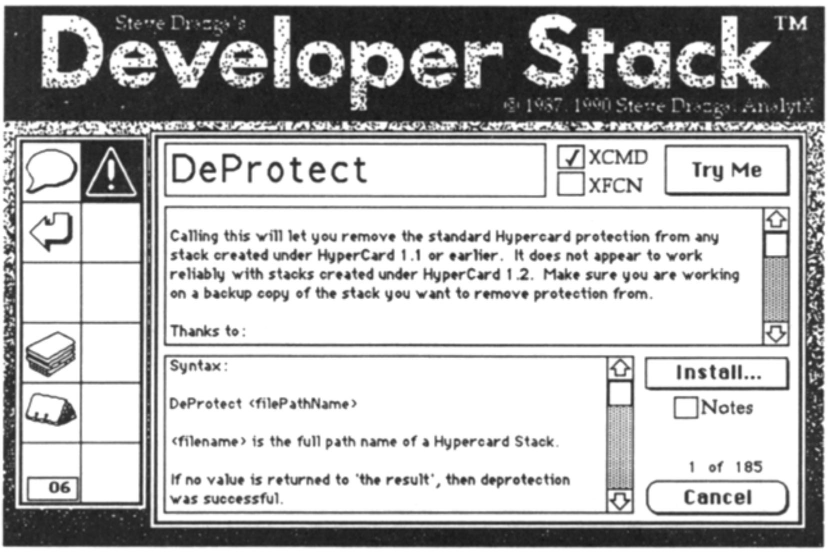


FIGURE 7 □ continued





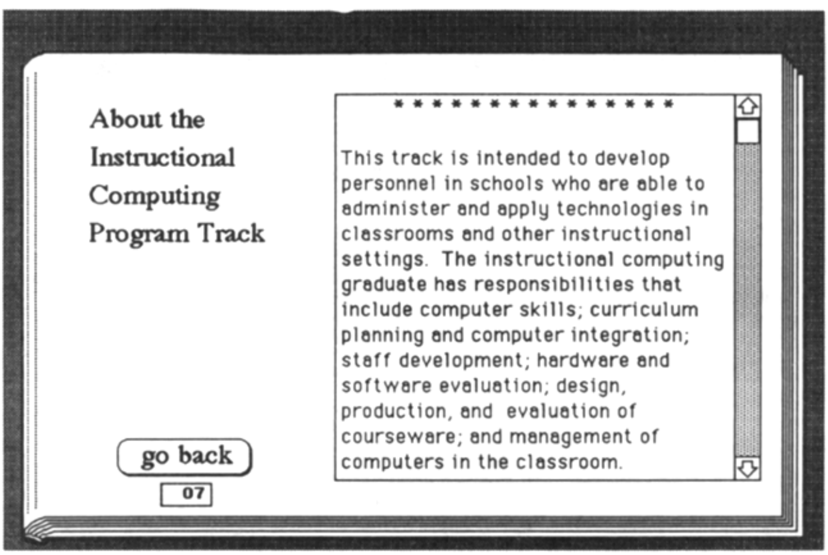
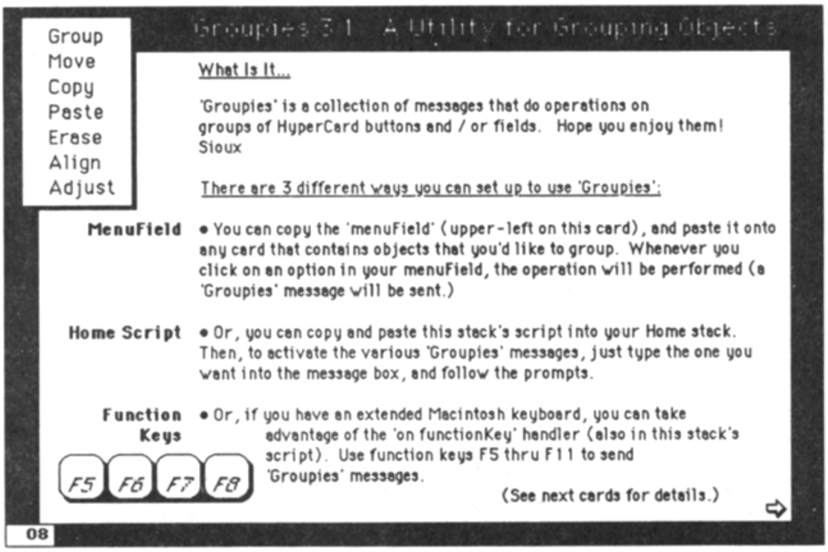
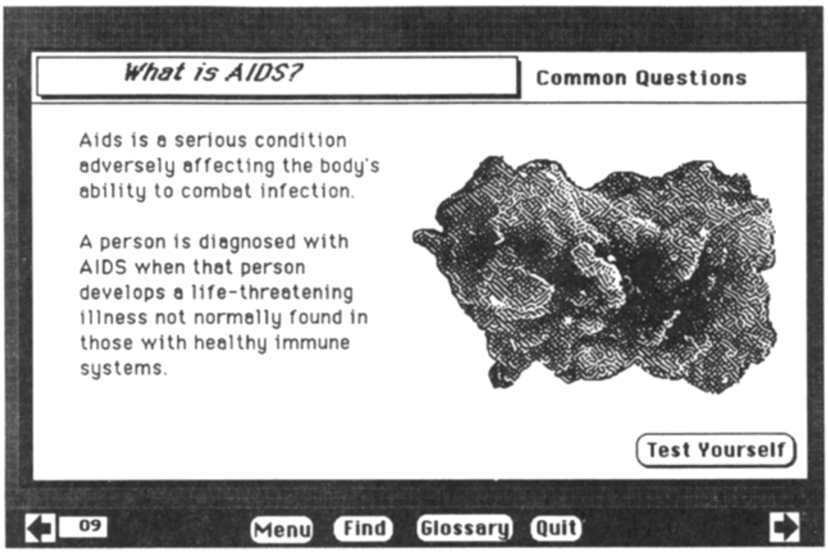


FIGURE 7 □ continued





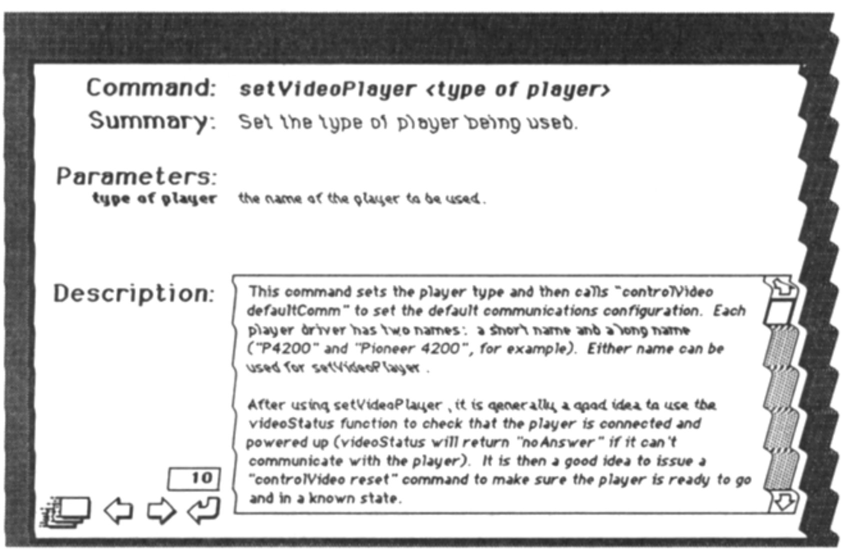
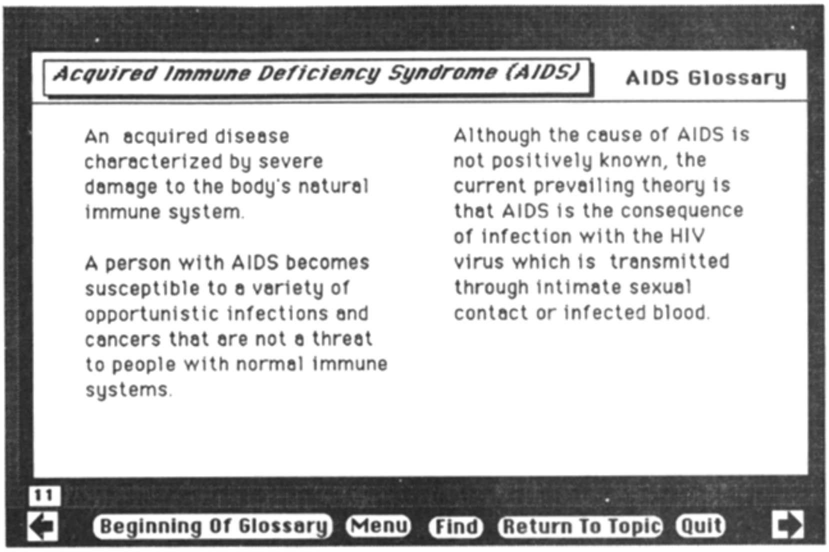
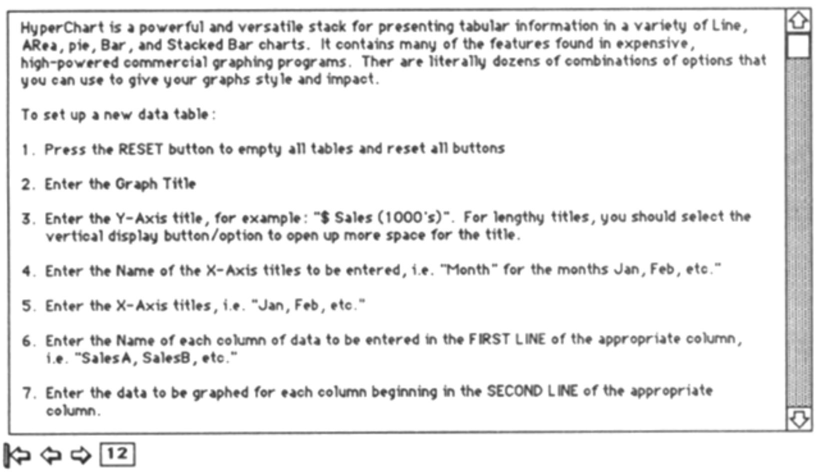


FIGURE 7 □ continued





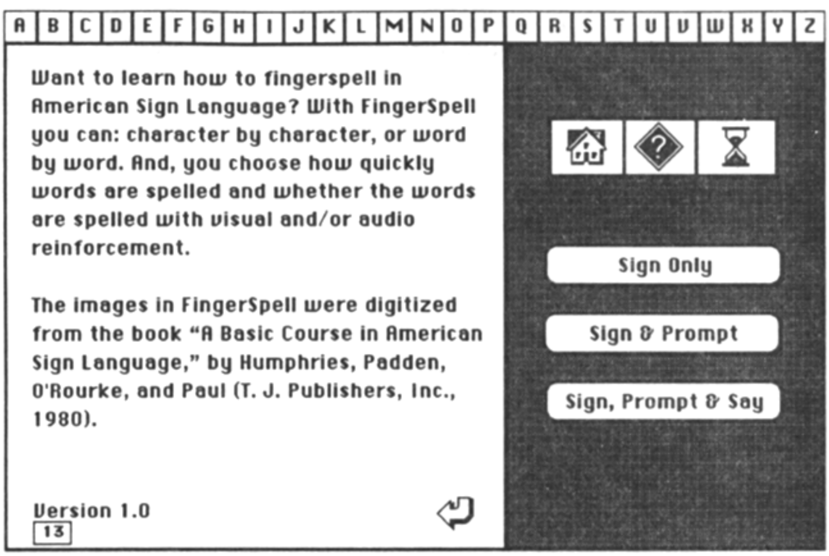
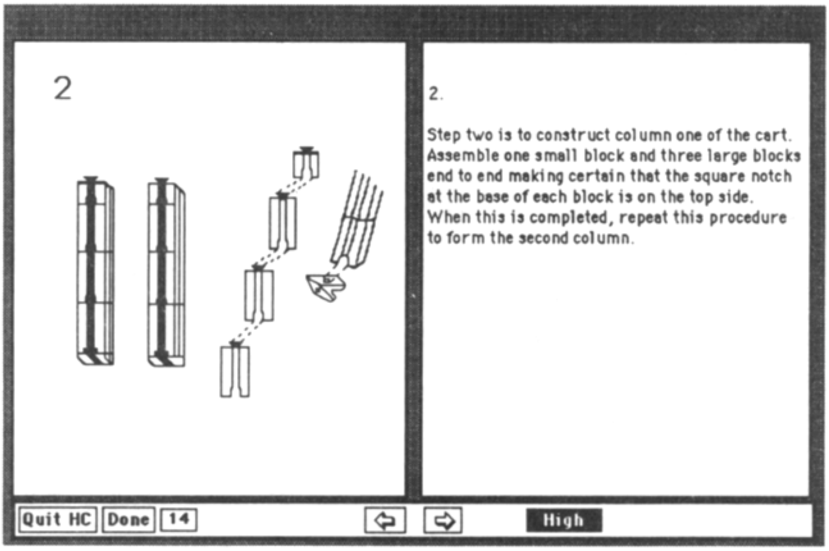


FIGURE 7 ☐ continued



MULTIDIMENSIONAL SCALING (or MDS) is a set of mathematical techniques that enable a researcher to uncover the "hidden structure" of data bases, as illustrated below. The authors, who are among the pioneers in developing and using these techniques, deal very concretely with the problems really faced in using them, and present varied applications.

An example illustrating an interesting MULTIDIMENSIONAL SCALING application in political science involves data from a 1968 election study conducted by the Survey Research Center of the University of Michigan. Each respondent in a national sample evaluated 12 actual or possible candidates for President of the United States. How similarly did the public view the candidates? What identifable features can we discern in the varying evaluations of the candidates that can help us understand what led individual citizens to their decisions? MULTIDIMENSIONAL SCALING can help answer these questions by locating the political candidates in a spatial configuration or "map." Once we have located the candidates or points in (multidimensional) space, we seek to determine the hidden structure, or theoretical meaning of this spatial reprresntation of candidates.

Applying MULTIDIMENSIONAL SCALING to these data provides a way of reducing the data about 12 candidates to two dimensions representing the

15

MULTIDIMENSIONAL SCALING (or MDS) is a set of mathematical techniques that enable a researcher to uncover the "hidden structure" of data bases, as illustrated below. The authors, who are among the pioneers in developing and using these techniques, deal very concretely with the problems really faced in using them, and present varied applications.

An example illustrating an interesting MULTIDIMENSIONAL SCALING application in political science involves data from a 1968 election study conducted by the Survey Research Center of the University of Michigan. Each respondent in a national sample evaluated 12 actual or possible candidates for President of the United States. How similarly did the public view the candidates? What identifable features can we discern in the varying evaluations of the candidates that can help us understand what led individual citizens to their decisions? MULTIDIMENSIONAL SCALING can help answer these questions by locating the political candidates in a spatial configuration or "map." Once we have located the candidates or points in (multidimensional) space, we seek to determine the hidden structure, or theoretical meaning of this spatial reprresntation of candidates.

Applying MULTIDIMENSIONAL SCALING to these data provides a way of reducing the data about 12 factors.

16

FIGURE 7 ☐ continued

Incomplete Matrices

Several procedures are available for obtaining estimates of the scale of values when, for one reason or another, the matrix "X" contains unfilled cells.

Traditional Procedure

If the unit of measurement is specified so that c is equal to unity, it follows that the theoretical equations for stimuls k and stimulus k+a can be written.

It is thus seen that, for errorless data, the difference for any value of j is equal to the difference in scale values. I like manner, the corresponding differences between observed x values is an estimate of theis difference in scale values. For any two stimuli, there will be as many such estimates as there are filled pairs of cells in the columns of matrix X.

Matrix P

	1	2	3	4
1		1.000	0.935	0.975
2	0.000		0.000	0.025
3	0.065	1.000		0.690
4	0.000	0.840	0.065	
5	0.025	0.945	0.310	0.160

Calculating Matrices

Theoretically, a in equation 29 may take any value from 1 to n-k. In actual practice, howerver, differences are obtained only for stimulit that are adjacent on the attribute being scaled. Adjacent stimuli will ordinarily have more filled cells in common and will give more reliable estimates of differences. The usual procedure when constructing the matrix X is to arrange its columns in rank order with respect ot the attribute. The rank is given by the rank order of the sums of the columns of matrix P.

17 | Home | Glossary | Contents | Help

The earth provided a view that exceed the moon's capacity to bore with its monochromatic color scheme.

Space: The Final Frontier

It was a lofty ethical struggle, all right, when I was offered an all-expenses-paid junket to Japan, courtesy of the Japanese Trade Ministry. Would I be compromised by the cusy first-class accommodations, and perhaps the stray geisha? After a deep examination of my soul which consumed somewhat less time than a font change on a 10-page Word file), I decided of course not!

Why did Jasmine quote different SIMM prices on the same day? Jasmine's supply of $39.95 SIMMs disappeared rapidly, so many readers of the January and February issues were quoted a higher price than that advertised. Another change is the breadth of tech support. "Jasmine tech support used to provide a lot tech support on non-Jasmine questions, such as questions about software INITs.

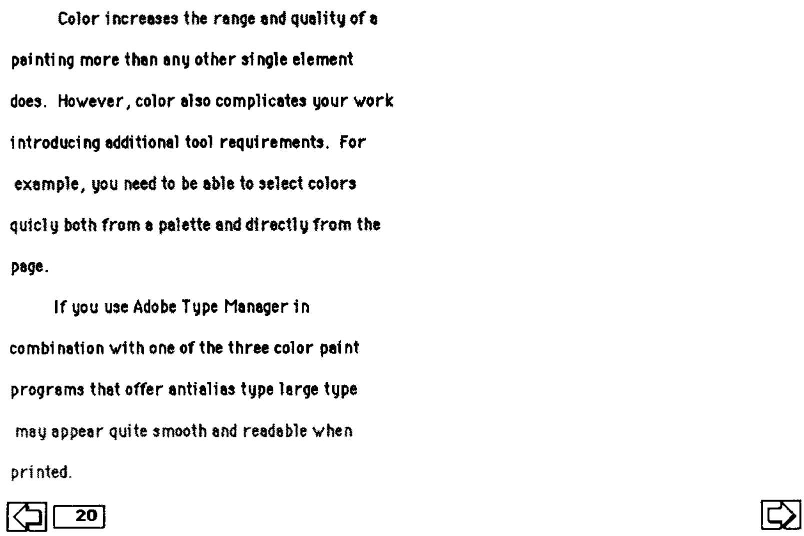
18

Color increases the range and quality of a painting more than any other single element does. However, color also complicates your work introducing additional tool requirements. For example, you need to be able to select colors quicly both from a palette and directly from the page.

If you use Adobe Type Manager in combination with one of the trhee color paint programs that offer antialias type large type may appear quite smooth and readable when printed.

19

FIGURE 7 ☐ continued



Results

Participants (N = 33) performed comparisons on 20 real screens. Results from the MDS analysis of judgment data indicate that the subjects evaluated the real screens along a single dimension. Figure 8 presents the stress results that occurred when the participant utilities were run through one-, two-, three-, and four-dimensional solutions. The MDS solution chosen for interpretation was limited to one dimension because the level of stress was extremely low (.009). Stress actually increased in the subsequent MDS analyses, indicating that the MDS procedures did not identify any other spatial arrangement that represented the similarities and differences among the screens. The resulting one-dimensional solution implies that there is one primary construct or criterion on which viewers judged the readability and studyability of the real screens.

Figure 9 presents the MDS results in a single one-dimensional figure. Recall that MDS results have no built-in procedures for labeling the dimension. The first step, then, in interpretating the MDS results for the real

FIGURE 8 ☐ MDS Stress Values for Real Screens along Four Dimensions

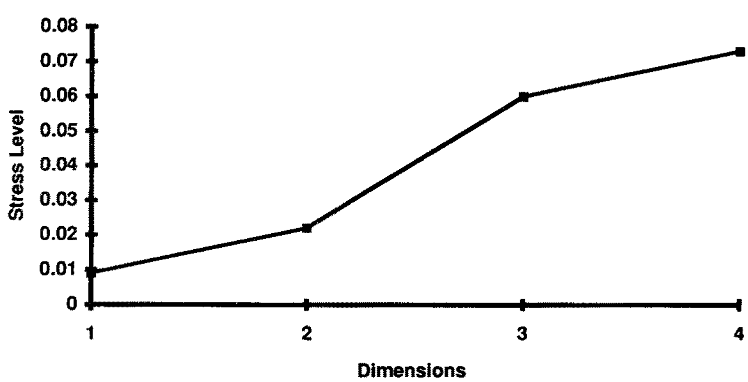
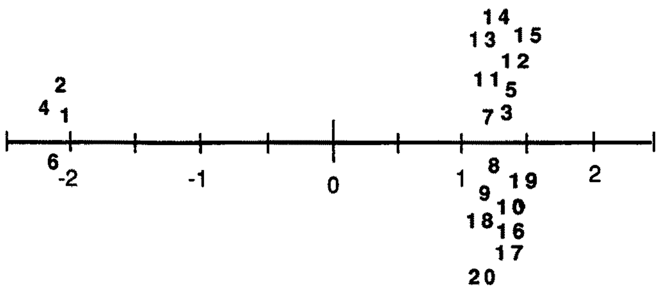


FIGURE 9 ☐ MDS Configuration for Real Screens in One Dimension



screens is to examine the stimuli locations on the map and axis of the dimension in terms of the objective, visible characteristics of the stimuli. The locations of the stimuli along the axis represent subjects' perceptions of how similar or different one screen is from the others in terms of readability and studyability. The actual coordinates for the real screens are found in the "Coordinate" column of Table 8. Note that the screens are grouped into two distinct groups: screens 1, 2, 4, and 6 and screens 3, 5, and 7–20.

Another step in the interpretation of the one-dimensional MDS solution for real screens is the ranking of the utility means of the stimuli from the raw rating scores to identify the most studyable and readable stimuli (see the columns labeled "rank" and "mean" in Table 8). Screens 1, 2, 4, and 6 have a more positive mean ranking (mean = 27.5) than the other screens. Therefore, the left end of the dimension represents the more positive aspect of the dimension. Most of the screens are clustered together along the right end of the dimension, which probably means they were judged slightly less readable and studyable (M = 34.5).

Note that screen 9, the highest rated screen, actually falls in the larger group. This is an example of a peculiarity of MDS techniques and the effect of less-than-perfect stress (perfect stress = .000). The groupings are a result of statistical compromise, and the MDS analysis resulted in screen 9 being placed in the larger group.

Finally, to help interpret the MDS dimension, adjectives with which the participants rated the screens were regressed on the MDS coordinates. In this case, participants rated each of the 20 screens immediately following the paired-comparison task. The same adjectives used in Study 1 (Figure 5) were used in this study. As in Study 1 (see Table 6, for example), the correlation matrix showed a high degree of intercorrelation among the adjectives. Therefore, a principal components factor analysis using an oblique rotation was conducted to create a more distinct set of descriptors. Table 9 shows the three factors. While Study 1 showed two factors, Study 2 added

TABLE 8 ☐ Real Screen Coordinates and Rank

Screen	*Coordinate*	*Rank*	*Mean*
SCR01	−2.11	8	28.7
SCR02	−2.10	3	22.2
SCR03	1.30	12	35.8
SCR04	−2.12	7	28.2
SCR05	1.30	15	38.5
SCR06	−2.12	11	30.7
SCR07	1.29	4	25.5
SCR08	1.30	13	36.7
SCR09	1.29	1	21.7
SCR10	1.30	14	29.5
SCR11	1.30	9	29.2
SCR12	1.31	2	21.8
SCR13	1.29	19	45.4
SCR14	1.30	10	29.5
SCR15	1.31	20	49.0
SCR16	1.31	18	45.1
SCR17	1.30	16	39.4
SCR18	1.29	5	26.7
SCR19	1.29	6	27.4
SCR20	1.31	17	43.3

Note: Low mean is judged most studyable and readable.

TABLE 9 □ Real Screen Ranks by Utility Mean

Rank	*Screen*	*Mean*	*Rank*	*Screen*	*Mean*
01.	SCR09	21.7	11.	SCR06	30.7
02.	SCR13	21.8	12.	SCR03	35.8
03.	SCR02	22.2	13.	SCR08	36.7
04.	SCR07	25.5	14.	SCR10	37.0
05.	SCR18	26.7	15.	SCR05	38.5
06.	SCR19	27.4	16.	SCR17	39.4
07.	SCR04	28.2	17.	SCR20	43.3
08.	SCR01	28.7	18.	SCR16	45.1
09.	SCR11	29.2	19.	SCR12	45.4
10.	SCR14	29.5	20.	SCR15	49.0

Note: Low mean is judged most studyable and readable.

one more factor. This indicates that there was a richer set of stimuli to describe, making the discriminations among the stimuli easier and richer.

The adjectives that loaded on Factor 1 (Table 10) were: "attractive," "readable," "studyable," "interesting," "inviting," and "dynamic." (Remember that the pattern loadings are used to interpret the unique contribution of each variable to that factor.) These adjectives refer to the aesthetic qualities of the screen and indicate that users evaluated screens on their visual interest; accordingly, this factor was labeled "interesting." The adjectives that loaded on Factor 2 were: "neat," "clean," and "organized." These adjectives refer to the use of space to create an organized display and were collectively labeled "organization." The adjectives that loaded on Factor 3 were: "planned," "structured," and "controlled." These adjectives seem to refer to the inherent structure of the text elements to indicate content structure; therefore, this factor was labeled "structure."

Next, the factor scores resulting from the factor analysis were regressed onto the dimension. The results of the multiple regression showed that Factor 1, interesting, loaded significantly on the dimension ($t = 4.46$, $p = .0001$). Factor 3, structure, also loaded significantly on the dimension ($t = 5.093$, $p = .0001$). Factor 2, organization, was not significant ($t = 1.006$, $p = .315$). With this information and the rankings and spatial map of MDS coordinates, the dimension can be interpreted.

MDS Dimension: Aesthetic and Organizational Qualities

On a visual basis, the MDS dimension (Figure 9) clearly separates screens 1, 2, 4, and 6 (most readable and studyable) from screens 3, 5, and 7–20 (least studyable and readable). The former group of screens along this dimension show strong aesthetic and organizational qualities. Screens 1, 2, 4, and 6 are aestheti-

TABLE 10 □ Oblique Factor Analysis of Adjectives for Real Screens

	PATTERN LOADING			STRUCTURE COEFFICIENTS		
Adjectives	*Factor 1*	*Factor 2*	*Factor 3*	*Factor 1*	*Factor 2*	*Factor 3*
1. Neat	−.56	1.24*	.12	−.26	.71	.06
2. Clean	−.72	1.29*	.20	−.36	.75	.11
3. Organized	.00	.90*	−.01	.00	.52	−.01
4. Spacious	.36	.78*	−.30	.17	.45	−.16
5. Attractive	.79*	.22	−.06	37	.13	−.03
6. Planned	.16	.16	.61*	.08	.09	.33
7. Readable	.56*	.35	.00	.26	.20	.00
8. Studyable	.77*	.33	−.17	.36	.19	−.09
9. Interesting	.98*	−.09	.00	.46	−.05	.00
10. Structured	−.40	−.25	1.36*	−.19	−.15	.73
11. Inviting	.95*	.16	−.21	.45	.09	−.11
12. Dynamic	1.05*	−.23	.00	.49	−.14	.00
13. Controlled	.03	−.06	.88*	.01	−.03	.47

*Adjective loads on this factor.
Note: The factor procedure used was principal components analysis using an oblique solution primary pattern matrix transformation method.

cally well-balanced with defined functional areas, multiple columns of text and buttons, and with white space that is around the exterior screen margins. The screens also use boxes, icons, shading, and illustrations. They are complex and interesting, but not too complex, with easily identifiable areas of text, buttons, and illustrations. The content appears structured using headings, some spaced paragraphs, and bold or italics for emphasis.

The other screens (3, 5, 7–20) show two general characteristics that are viewed negatively: they are either filled with text—plain, crowded, and with few graphics—or they are unbalanced and unattractive, with large areas of white space and a disorganized appearance. The adjective loadings also support this visual interpretation. Adjective Factors 1, interesting, and 3, structure, describe the dimension. Viewers separated the screens on this dimension by placing screens that are attractive, interesting, planned, and structured on the left end of the dimension.

Field Articulation, Conceptual Style, and Gender

The same MANOVA design used in Study 1 was used to compare field articulation, conceptual styles, and gender to the utility rankings of each screen. No differences were found, with the results being Pillais $F = 2.703$, $p = .128$ for field articulation; Pillais $F = 1.427$, $p = .860$ for analytical conceptual style; Pillais $F = 1.301$, $p = .954$ for relational conceptual style; and Pillais $F = .789$, $p = .212$ for gender.

DISCUSSION

The Dimensions: Criteria for Evaluating Readability and Studyability

The purpose of using multidimensional scaling techniques is to discover constructs that may help explain phenomena or to guide the design of instructional computer screens. In the case of model screens, the MDS techniques uncovered two evaluative dimensions. In the first dimension, viewers separated the screens by perceived organization. The definition of functional areas with space, boxes, and lines and the use of headings, directive cues, and spaced paragraphs combined to present a planned, controlled, organized, and structured appearance (see, for example, screens 1 and 5 in Figure 3). Organization was the most important dimension for these viewers because they were selecting screens from which to study. Perception, reading, and studying are processes that require organization of elements as an initial step (Fleming & Levie, 1993), so any design that employs techniques that result in a well-organized display appeals to viewers more than a design that employs few such techniques.

The second dimension from the model screens indicates that visual interest is also an important criterion. Screens that are plain, simple, unbalanced, and bare are perceived as undesirable (see, for example, screens 30 and 32 in Figure 3). Visual interest refers to screens that use text elements in such a way as to create an environment that invites exploration. A moderate degree of complexity is part of this environment, and the same factors that can help create organized screens contribute to a complex, visually interesting screen: lines, boxes, illustrations, and the placement of white space along the margins of the screen.

The results of Study 1, using a larger sample of model screens and eight controlled text elements, refine the definitions of earlier identified dimensions (Grabinger, 1984, 1987; Grabinger & Amedeo, 1988). However, this study differed from earlier studies in that only two criteria or dimensions were found rather than three. Specifically, one of the two dimensions identified here appeared to reflect a combination of two of the earlier dimensions: organization and structure. In addition, the earlier studies defined spaciousness, but in this study, visual interest seemed to be a more important criterion. These differences may be due to two reasons. First, the larger sampling of screens presented a great many options with little variability, so the vast majority of screens were perceived as similar and only a few stood out as unique (see horizontal axis of Figure 6). This lack of variation probably served to narrow the differences and hence the criteria used to evaluate those differences. Second, spaciousness was a tertiary criterion in the earlier studies and probably was not used by viewers in this study. The eight text elements, each with two values, created a larger number of options, with few of the screens exhibiting a significant amount of spaciousness. So, instead of separating screens by spaciousness, viewers used the criterion of visual interest to evaluate the more complex model screens of this study. In sum, this implies that when viewing model screens, viewers first look for screens that are organized to help them in the process of studying, and then look for screens that are intriguing or visually interesting.

While the use of models permits an examination of viewer perceptions in a controlled environment, generalizing from the model screens to real applications may not be tenable (Ross & Morrison, 1989). The purpose of comparing judgments of real screens against model screens was to determine if the same results carried across to more practical applications. Study 2 indicated, as did Study 1, that organization and visual interest are important criteria in judging the readability and studyability of the real screens. For example, screens 1, 2, 4, and 6 (see Figure 7) present displays that are complex yet well organized, with distinct areas for buttons, illustrations, text, and titles. These screens are visually complex and interesting and give the impression that the content is well planned and well organized.

While these results are quite similar to the results of the model screens, it is surprising that there were no secondary or tertiary criteria or dimensions. It seems that both of the dimensions found in the model screens collapsed into a single criterion in Study 2. The real screens show a great deal more variety than the model screens, which seemed to facilitate making a decision. The model screens were judged very similar and required more careful analysis and the construction of alternative criteria for making judgments. Finally, the actual content of the real screens seemed to have no effect on judgment. There seemed to be no common theme among the four screens (1, 2, 4, and 6 in Figure 7), nor did informal post-study interviews indicate that content was a factor in judging the screens.

Individual Differences

Neither study found any generalizable differences or preferences as a factor of differences in field articulation, gender, or conceptual style of viewers. The few significant comparisons appeared to be isolated findings without clear meaning or support from related analyses. Admittedly, this exploratory effort was to determine if those individual differences mattered in judging readability and studyability from the appearance of the screens. Since the task examined the most initial phases of perception and did not require extended analysis of the parts of the screens, there probably was not time for field articulation and conceptual style to play a significant role. One difference that was not analyzed but that may be a factor is familiarity with the Windows or Macintosh operating environments. Subjects who have familiarity and experience with these environments may have developed expectations about what a screen should look like, while novices may not have such preconceptions.

Design Recommendations

The constructs of organization and visual interest provide some rules of thumb for arranging numerous text elements to create readable and studyable screens. Rather than focus on individual text elements, computer-based instruction and hypermedia producers can instead focus on arranging text elements so as to create organized, structured, and visually interesting screens.

Provide a macro level of organization. Generally, the most useful way to operationalize these constructs is to organize the screen into functional areas. Designers should decide on where status and progress information, navigation buttons, content displays, control buttons, and illustrations will be located, and use graphic devices such as shading, lines, and boxes to separate one area from another. Screens 1, 2, 4, and 9 in Figure 7 are good examples: titles, text, and buttons are clearly separated using lines, white space, and shading. This design technique works only when consistency is also practiced. The functional areas should appear in the same locations, and the devices used to define them should be the same throughout a program and its parts.

Use structure to create a micro level of organization. Following these macro-level analyses, designers should then consider how the screen can reflect the structure of the content. Generally, users prefer screens that use headings, directive cues, and spaced paragraphs to indicate the hierarchy of the content and to break the content into studyable chunks of information (see screens 8, 9, 10, and 17 in Figure 7). This can be achieved by using headings as organizers and directive cues to point out important terms and phrases, by using increased spacing between paragraphs rather than traditional indentation, and by showing comparisons in side-by-side columnar arrangements.

Provide visual interest. Finally, designers should consider the visual interest of the screen. Viewers dislike screens that are plain or full of text without any headings, directive cues, lines, shading, buttons, titles, or illustrations (such as shown in screens 12, 15, and 16 in Figure 7). It seems that a variety of well-organized text elements enrich the environment and make it more interesting to explore. Excessive complexity results when too many elements or too much information is crammed on the screen.

Research Recommendations

Although the population and stimuli samples used in this study limit generalizability, in view of consistency with past studies, the findings provide a base for guiding the design of screens. While the dimensions discussed here are interpretations and thus open to further explanation and refinement, programs could be designed with screens to reflect the use of the constructs of organization (at both the macro and micro levels) and visual interest to study effects on learning, motivation, and the creation of knowledge structures.

However, evaluation in this field of study is problematic and needs creative solutions. Paper-and-pencil recall and essay tests may not be the best measure of how people use screen displays to construct knowledge structures, nor are they always the best way to measure problem-solving ability, creativity, or cognitive flexibility. It may be possible to compare a program with highly organized screens to a program with plain screens by asking learners to create an outline of the material just read to see which program facilitates the creation of an outline. Another possibility is to use semantic map-generation programs to assess how the structure of a set of screens affects the semantic meanings generated while students study the program. In combination with essay and problem-solving activities, such an assessment could determine whether outlines and semantic maps facilitate the development of flexible, creative problem-solving skills in new contexts.

While the concepts of organization and visual interest may seem to be common sense, common sense is not always followed by designers. In addition, knowledge of these concepts does not mean that designers will apply them in ways that create aesthetically pleasing and well-organized screens. The present study suggests that these constructs *are* important to prospective viewers and may help gain attention and build confidence in using instructional material (Keller & Burkman, 1993). □

R. Scott Grabinger is an Associate Professor with the Division of Instructional Technology at the University of Colorado at Denver.

Special thanks are due to Barbara Grabowski and Larry Ragan at Pennsylvania State University for aiding in the administration of the studies, and to Ernest Rakow at Memphis State University for providing helpful advice on statistical analyses.

REFERENCES

Bovy, R. C. (1981) Successful instructional methods: A cognitive information processing approach. *Educational Communications and Technology Journal, 29*(4), 203–217.

Brainpower, Inc. (1986). Statview 512+. Calabasas, CA: Author.

Donahue, B. (1978). *The language of layout.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Educational Testing Service. (1975). *Hidden Figures Test*. Princeton, NJ: Author.

Fleming, M., & Levie, W. H. (Eds.). (1993). *Instructional message design: Principles from the behavioral and cognitive sciences* (2nd ed.). Englewood Cliffs, NJ: Educational Technology Publications.

Garrett, L. (1967). *Visual design: A problem-solving approach*. New York: Van Nostrand Reinhold.

Grabinger, R. S. (1984). CRT text design: Psychological attributes underlying the evaluation of models of CRT text displays. *Journal of Visual Verbal Languaging*, 4(1).

Grabinger, R. S. (1987, winter). Relationships among text format variables in computer-generated text. *Journal of Visual Verbal Languaging*. 6(3).

Grabinger, R. S. (1989). Reconsidering CBI screen design: Alternative to text-based designs. *Journal of Computers and Human Behavior*, 5(3), 175–183.

Grabinger, R. S., & Amedeo, D. (1988). CRT text layout: Perceptions of viewers. *Journal of Computers and Human Behavior*, 4, 189–205.

Green, P. E. (1974, September). On the design of choice experiments involving multifactor alternatives. *Journal of Consumer Research*, 1, 61–68.

Hair, J. F., Anderson, R. E., Tathman, R. L., & Grablowsky, B. J. (1979). *Multivariate data analysis with readings*. Tulsa, OK: Petroleum Publishing Co.

Hartley, J. (1985). *Designing instructional text*. London: Kogan Page.

Kagen, J., Moss, H. A., & Siegel, L. E. (1963). Psychological significance of styles of conceptualization. *Monographs of the Society for Research in Child Development*, 28(2), 73–112.

Keller, J., & Burkman, E. (1993). Motivation principles. In M. Fleming & W. H. Levie, *Instructional message design: Principles from the behavioral and cognitive sciences*. Englewood Cliffs, NJ: Educational Technology Publications.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills, CA: Sage Publications.

Misanchuk, E. R. (1992). *Preparing instructional text: Document design using desktop publishing*. Englewood Cliffs, NJ: Educational Technology Publications.

Nelson, R. P. (1978). *Publication design* (2nd ed.). Dubuque, IA: Wm. C. Brown.

Rehe, R. E. (1979). *Typography: How to make it most legible* (3rd ed.). Carmel, IN: Design Research International.

Reynolds, L. (1979). Legibility studies: Their relevance to present day documentation methods. *Journal of Documentation*, 35(4), 307–340.

Ross, S. M., & Morrison, G. R. (1989). In search of a happy medium in instructional technology research: Issues concerning external validity, media replications, and learner control. *Educational Technology Research and Development*, 37(1), 19–33.

Rummel, R. J. (1970). *Applied factor analysis*. Evanston, IL: Northwestern University Press.

Schiffman, S. S., Reynolds, M. L., & Young, F. W. (1981). *Introduction to multidimensional scaling: Theory, methods, and applications*. New York: Academic Press.

Spence, I., & Domoney, D. W. (1974). Single subject incomplete designs for nonmetric multidimensional scaling. *Psychometrika*, 39(4), 469–489.

SPSS, Inc. (1988). *SPSS-X user's guide* (3rd ed.). Chicago, IL: Author.

Tinker, M. A. (1963). *Legibility of print*. Ames, IA: Iowa State University.

Tinker, M. A., & McCullough, C. M. (1962). *Teaching elementary reading* (2nd ed.). New York: Appleton-Century-Crofts.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.

Turnbull, A. T., & Baird, R. N. (1975). *The graphics of communication: Typography, layout, design* (3rd ed.). New York: Holt, Rinehart and Winston.

Twyman, M. (1981). Typography without words. *Visible Language*, 15(1), 5–12.

Waller, R. H. W. (1980). Graphic aspects of complex texts: Typography as macropunctuation. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language* (Vol. 2). New York: Plenum Press.