Learner/User Preferences for Fonts in Microcomputer Screen Displays

Earl R. Misanchuk

ABSTRACT: In a forced-choice comparison mode, subjects were asked to judge which of two screen displays using different fonts on a Macintosh microcomputer were the easiest to read and to study from. Identical "dummy" text (except for font) was used for all screen displays, Twelve-point Courier, Monaco, Geneva, Boston, New York and Chicago were compared. All screen displays were presented, and data were collected, via a HyperCard stack designed for the purpose.
Sixty-two subjects determined that Geneva was the preferred font, with Boston a reasonably close second choice. Chicago was the last choice, with Courier not much more popular. Monaco and New York were somewhat below the midpoint of the top and bottom groups.

As presentation technology evolves, not only must new questions be addressed, but questions which have already been answered more-or-less satisfactorily through research must be re-examined periodically to determine if the findings are still relevant. A good example of this maxim is in the field of text displays on computer screens. Hartley (1987), in a recent comprehensive review of the field, cites a number of research studies dealing with typefaces on computer screen displays (Maddox, Burnette, & Gutmann, 1977; Riley & Barbato, 1978; Snyder & Taylor, 1979) as "...research...being carried out to see what fonts seem to be the most legible and most preferred" (Hartley, 1987, p. 8). Close examination of these studies shows that they actually deal with single-stroke, boxy, 5x7 dot-matrix fonts, the likes of which are rarely used any more on modern computer screens. The time lag involved in publishing research and deriving generalizations from published studies is such that it is not uncommon that a generalization is obsolete as soon as it is made.

Even the most recent of the studies Hartley cites was done prior to the advent of the Apple® Macintosh™ microcomputer. The Macintosh (or Mac) is rapidly making inroads into educational and training environments because of its versatility and ease of use, and its ability to integrate text with graphics readily. A standard feature of the Mac (and increasingly, of competing computers as well) is its ability to employ several different typefaces (or fonts, as they

are known in Mac parlance), styles (e.g., italic, bold, underlined, etc.), and sizes within a single screen display - a rather dramatic change from microcomputers preceding the Mac, and many extant today. Most application programs designed for the Mac permit the user to choose from among a number of fonts, allowing the instructional designer to use these choices as design elements.

The assumption of an 80 character x 24 line computer screen display is no longer valid. With the Mac's ability to change fonts, size, and style, and to use either proportionally-spaced or mono-spaced type, some of the conclusions made by researchers such as Maddox, et al. (1977), Riley and Barbato (1978), and Snyder and Taylor (1979) - although only a decade old - may not necessarily be relevant any longer.

Yet generalizations about which fonts to use on a computer screen are obviously important to the instructional designer intending to use that medium. Along with the increased dissemination of microcomputers comes the provision to the multitude of their users powerful programs that were, until very recently, only available in a mainframe environment. Two examples are computer assisted instruction (CAI) authoring systems such as Authorware's Course of Action and Best Course of Action, and hypermedia such as OWL's Guide and Apple's HyperCard. As these very powerful tools come into increasingly common use, an increasing public needs to become aware of screen design considerations (e.g., what font to use for maximum effect).

The question of what font to use in print media has been well addressed, and generalizations are available for designers working in that medium. Dreyfus (1985), in speaking of printed text, states:

> The outcome of many experiments indicates that there is no statistically significant difference between the legibility of a wide variety of text types, even between seriffed and unseriffed types. On the other hand, differences of real statistical significance were detected when readers were asked which styles of type they preferred.. .This finding ought to be studied by those who decide in what types to compose the vast amount of printed matter that is intended to attract or to persuade, but which nobody is obliged to read. (p. 18)
> .. .a great deal of attention ought to be paid to the peculiar problems of devising type designs tailored to the changed conditions under which so much knowledge is now transmitted, (p. 22)

It is reasonable to assume that similar findings will obtain with regard to computer screen text. If none of the commonly-used fonts displays a readability advantage over another, then designers might look to learner/user satisfaction as a guide for choosing a font, on the theory that increased satisfaction should maximize learner/users' motivation and attention. Hooper and Hannafin (1986) put it this way:

It may well be that the measurable effect of each of the variables [related to computer-generated text] on learning is minimal. However, the overall effect of reading text from a screen that is pleasant to look at may in itself have positive transfer to learning. Designers of computer based instruction are virtually unaffected by cost limitations when organizing text display. Consequently, the potential impact of different modes of presentation maybe considered, without fear of increasing production cost, while possibly capturing the readers' attention and helping to organize information. This may result in text that is both easier to read and better organized in long term memory, (p. 27)

Whether the thrust of communication is instructional, informational, or even (perhaps especially) persuasive, the affective characteristics of the elements comprising the communication cannot be ignored.

This study undertook to determine differences among a number of fonts available to the instructional designer working with a Macintosh, and particularly addressed the question of learner/user preference of font.

Fonts Compared

Six fonts commonly found on Apple® Macintosh™ microcomputers were chosen for comparison: Chicago, Geneva, New York, Monaco, Courier, and Boston (see Figure 1 for samples). These particular fonts were selected from the many hundreds of fonts available for the Macintosh primarily on two bases: their widespread use, and their having been designed as Macintosh screen fonts (as distinct from Macintosh LaserWriter™ fonts, which show up relatively poorly on the screen). In all cases, the 12-point size font was employed.

Figure 1.
Samples of Fonts Used (actual size). The discontinuity evidenced in italic styles is an artifact of their reproduction on paper; they appear to the eye to be more continuous when on the screen.

This is a sample of Chicago. It can also be **bold** or *italic,*

This is a sample of Geneva. It can also be bold or *italic*

This is a sample of New York. It can also be **bold** or *italic*

This is *a* sample of Monaco.   It can also be **bold** or *italic,*

This is a sample of Courier.   It can also be **bold** or *italic.*

This is a sample of Boston. It can also be **bold** or *italic*

Chicago is the font used by the Macintosh operating system for menus and buttons. It is therefore both familiar and guaranteed to be found on all Macintosh computers - indeed, it is built into the ROM's of the Mac 512KB, Mac Plus, Mac SE, and Mac II (Poole, 1988). Geneva is a commonly used default font for many applications programs, and is also in the SE and Mac II ROM's (Poole, 1988). Geneva is a sans-serif font that, along with New York, Monaco, and later Courier, was one of the fonts that were provided by Apple as part of all Macintosh operating systems. New York is also commonly used, since it is an original serif font that looks good when printed on the ImageWriter™. Monaco 9 is a font used by the operating system, and since Macintosh users frequently keep families of fonts on their system files, it was judged likely that most Macintosh users would have Monaco 12 available as well. Courier was included because of its similarity to the typeface of the same name popularized by the IBM* "Selectric" typewriter element and subsequent widespread use on daisy wheel printers and laser printers. Boston is a font that is more recent than the others, but has gained widespread use since it was bundled with Microsoft™ Word 3.0x.

USE OF "DUMMY" TEXT

The choice of what to put onto the screens is not trivial. To ensure that the content of the text does not affect subjects' perceptions, it seems important to maintain the long tradition in social science research of using nonsensical stimuli. Doing so, however, presents its own risks to validity.

In studies of text format, Grabinger (1984, 1985) employed a system of notation developed by Twyman (1981), in which

> ..."X"s were used to represent the bulk of the print on a page; "O"s to
> reflect the occurrences of italics, upper case, bold type, color, headings,
> or reverse type; and "I"s as a tertiary graphic unit to represent
> something particularly unique in style (Grabinger 1985, p. 4).

Twyman appears to have developed the notation system as a focal point for discussions about layout of text on a page, with students of typography and design. As such, it would probably be a useful tool to employ as a shorthand when making comparative statements about different page layouts. However, to expect the average person to be able to imagine the replacement of X's with "regular text", O's with any of a variety of specialized typefaces, and I's with some (undefined) "tertiary graphic unit…unique in style", may be asking too much.

Furthermore, judging from the examples published by both Twyman (1981) and Grabinger (1984), no attempt was made to represent individual words - entire lines of X's were used to represent the body of the text, and no punctuation was used. Grabinger's .j.985) study improved on the situation

somewhat by using groups of varying numbers of X's, some of which were printed in bold face, to represent words. Still, envisioning the resulting stimuli as bodies of actual text appears to involve a good deal of imagination.

This study attempted to define something half-way between the extremely abstract generalizations developed by Twyman and the possibly-too-concrete use of actual text. As is frequently done in layout and design of sample print products, "dummy" text was used.

An actual sample of text, including use of both italics and bold type, was transformed using the search and replace feature in a word processor. Thus all instances of the letter 'a' were replaced with T, 'b' with 'j', 's' with V, 'g" with 'a', and so on, to produce nonsensical text that resembled real text in syntactic structure and in word and sentence length (see Figure 2). By virtue of the Macintosh's ability to produce both italic and bold text on the screen, further verisimilitude was possible.

Figure 2.
Sample Screen Display (actual size), Using Geneva Font. Learner/users manipulated the mouse to point the finger to the appropriate "button."

Lidijice cgmdgibe pedhc ecebc dg cbeide cecdge lhhdlcidignc gidpged *Mgotiieeina* Dpic cpihdeb jeoinc gidp i cdeh-jy-cdeh "gidcdpbgeop" gm pgg i lidijice hiccioe Ic ecel dg cbeide in ihhdicidign mgb ceehino denind midec.

I lidijice ic i cgddecdign gm bedidel midec, ic cpggn in Mioebe 7-1. I mide cgncicdc gm becgblc dpid pgdl lidi.

Eidpeb i mide einioeb gb lidijice einioeeend cycdee hiccioe cgedl je ecel mgb dpe tfisw//ft/ihhd1cfdign leccbijel in dpe mlbcd hibd gm dpic cpihdeb. Ciehde hiccioec ibe cidd **mide einioebc.** Dpey ggbc gn gne **mide** id i diee. Mincieb hiccioec ibe ciddel **lidijice einioeeend cycdeec (LJEC).** Dpey cin ggbc gn cefebid midec id gne diee.

Dpe cpihdeb lieccbijec gn dpe denind mide jeoinc ic i mide einioeb ihhdicidign inl obggc indg i lidijice einioeeend cycdee ihhdicidign. Id idcg cpggc in ehiehged gm i LJEC hbgobie dinioeioe.

/•//<#? *einicietiCQb* **mide einioeeend hiccioec** ggbc gn gndy gne mide id i

Show me the other one   **1 choose this one.**

DATA-COLLECTION METHOD

All stimuli were presented and data were collected via a HyperCard™ stack. HyperCard is a program available solely on the Macintosh line of microcomputers, and uses as its metaphor a stack of cards (screen displays), each of which can be accessed in a variety of ways determined by either the author of the stack or the user of the stack. In this case, the author of the stack maintained control of how the stack could be used; only the pacing (speed of

response) was under the control of the user/subject. Cards (screen displays) were presented sequentially whenever subjects clicked the mouse on a particular area of the screen (these sensitized areas of the screen are referred to as "buttons").

The first four cards of the stack explained the purpose of the experiment and how the subject was expected to act in order to provide data. (The full text of the introduction is available from the author, as, indeed, is the HyperCard stack itself.) In effect, subjects were asked to choose which of two screen displays they thought would be the easiest to read and to study from.

Subjects moved to the next card by clicking on a designated button. On all of the introductory cards except the first, the subject also had the option of moving back one screen display to re-read portions of the introduction and instructions.

The next card asked for the subjects' first names and last names to be input, thanked them when the names were provided, and offered subjects the opportunity to either proceed or quit. Although the option existed for subjects to click on a "Cancel" button rather than provide either or both names, all subjects did provide their names and, when presented with the option to quit, all subjects proceeded.

The next two cards were sample comparisons, reiterating instructions on how to view the other of the pair of screens and how to indicate which screen was selected by the subject as the "best", and affording the subject an opportunity to get comfortable with the process of comparing the two screens making up the pair before actually being placed into a decision-making situation.

On the next 30 cards, the 15 pairs of screens representing the actual paired comparisons were presented (see Figure 2 for a typical screen display). As the sample comparisons and the actual comparisons were presented on the screen, two buttons appeared in an area at the bottom of the screen. Clicking on one of them brought onto the screen the other display of the pair under consideration. Clicking on the other one indicated that the screen currently displayed was the one that the subject had determined to be the one thought to be the easiest to read and to study from. It also caused the choice to be recorded, and the first screen of the next pair to be brought onto the screen.

The 15 pairs were presented in the order recommended by Ross (1934), such that a) the sequence of stimuli has no perceptible pattern of "correct" responses, b) pairs having a stimulus in common are maximally separated in the presentation order, and c) stimuli are balanced with respect to their order of presentation (so that no stimulus gets presented as the first of a pair more frequently than any other stimulus) (Torgerson, 1958, p. 168).

Finally, two cards asked subjects to indicate by "checking in a box" (using the mouse, of course) their level of education (some elementary school; completed elementary school; some high school; completed high school; some post-secondary education; completed a post-secondary degree, diploma, or certificate; completed more than ci^e post-secondary degree, diploma, or

certificate) and their age groups (under 10,11-20,21-30,31-40,41-50,51-60, 61-70, over 70).

The HyperCard stack was designed in such a way that as data were entered, they were saved onto a disk file, ready for analysis. This feature ensured that subjects' responses were not vulnerable to clerical error.

SUBJECTS

Ideally, subjects for a study such as this would comprise a random sample of the whole population with respect to all attributes (since it is not presently known which, if any, attributes may affect preference for certain fonts). In the light of there being no reason to believe that any particular attribute would bias the results, a non-random (but also generally non-systematic) sample was considered an acceptable substitute -in other words, a collection of people that share no apparent attribute that might affect their preferences for certain fonts. The only attribute shared by the subjects in this study was that they were the researcher's co-workers, family members, friends, and acquaintances. Otherwise, they had little in common: They varied in age and educational background, and in their experience with the Macintosh (see Figure 3 and Tables 1 and 2). Sixty-two subjects were involved in the study.
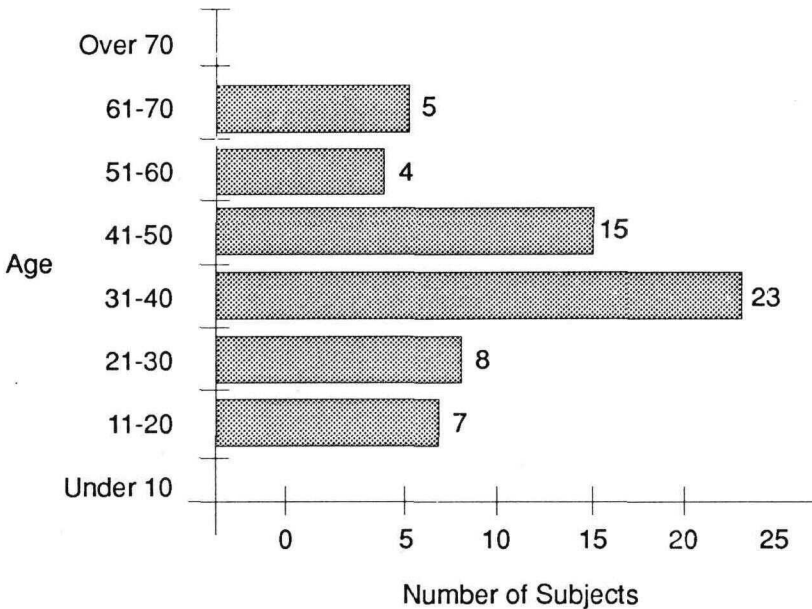
Figure 3.
Age Categories of Subjects.

TABLE 1
Highest Level of Education Attained by Subjects

Highest level of education attained	n
Some elementary school	0
Completed elementary school	0
Some high school	3
Completed high school	6
Some post-secondary education	11
Completed a post-secondary degree, diploma, or certificate	15
Completed more than one post-secondary degree, diploma, or certificate	27

TABLE 2
Subjects' Experience With Using the Macintosh

Experience	n
No experience or very limited experience	30
Some experience	24
Good deal of experience	8

RESULTS

The paired comparisons data were used to derive a Thurstone scale (Torgerson, 1958). The proportions matrix P that was generated is shown in Figure 4. Entries in the matrix show the proportion of times that the font comprising the column was chosen over the font comprising the row.

Figure 4.
*Proportions Matrix for All Subjects.*

	Courier	Monaco	Geneva	Boston	New York	Chicago
Courier	—	.69	.71	.74	.48	.44
Monaco	.31	—	.69	.55	.53	.39
Geneva	.29	.31	—	.48	.26	.23
Boston	.26	.45	.52	—	.24	.35
New York	.52	.47	.69	.76	—	.34
Chicago	.56	.61	.77	.65	.66	—

The Thurstone scale points, and their corresponding (easier-to-read) linear transformations, are shown in Table 3. The scale is shown in graphic form in Figure 5.

TABLE 3
Thurstone Scale Points

	Scale Point	Transformed Scale Point
Courier	-0.0369	-3.7
Monaco	-0.0106	-1.1
Geneva	0.0442	4.4
Boston	0.0290	2.9
New York	-0.0203	-2.0
Chicago	-0.0456	-4.6

Figure 5.
Thurstone Scale for All Subjects.
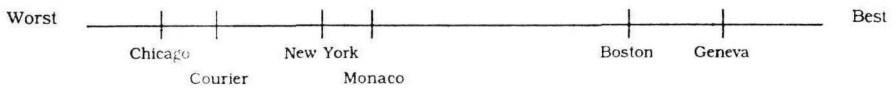
A useful feature of a Thurstone Scale is its ability to represent distances meaningfully. Thus, in the visual representation, it is quite easy to see which font was judged best, and by how much, relative to the others. Clearly, the font chosen as most easy to read and to study from by most people was Geneva. Boston was in second place, and relatively close to Geneva. Quite some distance behind were New York and Monaco, which were not very far apart on the scale. Courier and Chicago were judged not very different from one another, and both were a considerable distance behind the front-runners.

RELIABILITY

It was possible to "re-test" 28 of the original 62 subjects two weeks or more after their original "test". Because of vacations and other scheduling problems, the actual time between administrations varied somewhat from two weeks: Thirteen of the 28 subjects were re-tested exactly on time (two weeks later),

four were re-tested one day early, and all but one of the rest were re-tested from one to six days late. One person was re-tested 12 days late.

The number of times a subject's second judgment about a particular pair of stimuli agreed with the first judgment was calculated. The average number of agreements between the first and second data collections was 10.8 out of 15, with a standard deviation of 2.5 (see Figure 6).

Figure 6.
Number of Agreements Between Initial Responses and Responses Two Weeks or More Later.
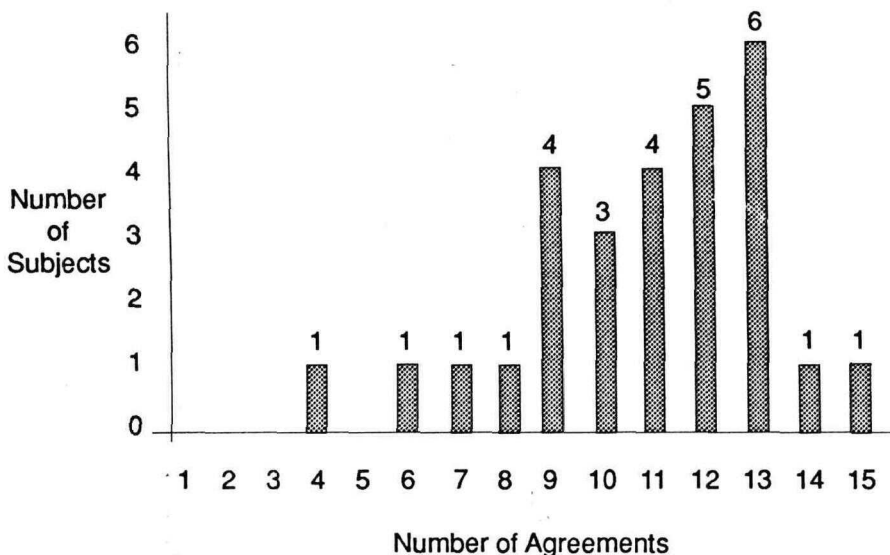
AGE AS A FACTOR

That eyesight generally becomes poorer with age is well known, and quality of eyesight might affect font choice. In order to get some idea whether there was an age effect operating, a median split was done on the sample with respect to age category. The responses of the 31 subjects in each of the "Younger" and "Older" groups were then used to construct separate Thurstone scales.

The proportions matrices for the two groups are shown in Figure 7 (see following page) and the corresponding graphic scales are shown in Figure 8. For the Younger group, Geneva was the preferred font, with Monaco and Boston (which were quite close together) some distance behind. Well behind the first three came New York, Courier, and Chicago (all quite close together). For the Older group, Geneva was again the front-runner, but Boston was a very

close second. Courier, Chicago, and Monaco formed a fairly tight cluster near the opposite end of the scale, with New York roughly midway between the two at the top of the scale and the three at the bottom.

Figure 7.
Proportions Matrices for Groups Formed by Median Split on Age.

a) Younger Group

	Courier	Monaco	Geneva	Boston	New York	Chicago
Courier	—	.81	.74	.74	.42	.45
Monaco	.19	—	.58	.35	.39	.32
Geneva	.26	.42	—	.42	.23	.19
Boston	.26	.65	.58	—	.19	.32
New York	.58	.61	.77	.81	—	.35
Chicago	.55	.68	.81	.68	.65	—

b) Older Group

	Courier	Monaco	Geneva	Boston	New York	Chicago
Courier	—	.58	.68	.74	.55	.42
Monaco	.42	—	.81	.74	.68	.45
Geneva	.32	.19	—	.55	.29	.26
Boston	.26	.26	.45	—	.29	.39
New York	.45	.32	.71	.71	—	.32
Chicago	.58	.55	.74	.61	.68	—

No tests of significance are available for Thurstone scale analysis, and the scale establishes only relative (not absolute) positions for stimuli, although the scale values are interval in character. Still, looking at the graphic scales, one might hypothesize an age effect: Monaco, which is one of the worst fonts as far as the Older group is concerned, ends up in a near-tie for second place as far as the Younger group is concerned. This age effect would have to be established more specifically in an experiment designed to focus on that variable, however, before definitive statements could be made. Indeed, perhaps the distinction is not worth making. Since Geneva ended up in first place with both groups, and Chicago and Courier ended up near the bottom of the scale for both groups, perhaps that is sufficient guidance for the instructional designer of HyperCard or other Macintosh screen displays.

Figure 8.
Thurstone Scales for Groups Formed by Median Split on Age.

(a) Younger Group.

(b) Older Group
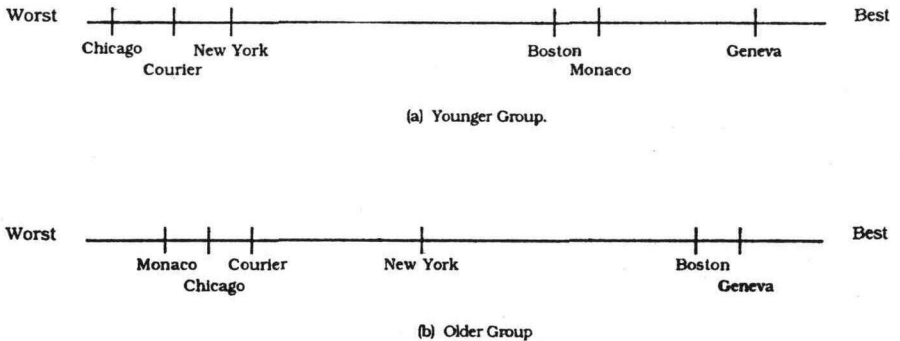
CONCLUSIONS

In a forced-choice comparison mode, subjects were asked to judge which of two screen displays using different fonts on a Macintosh microcomputer was the easiest to read and to study from. Identical "dummy" text was used for all screen displays (except for font). Twelve-point Courier, Monaco, Geneva, Boston, New York, and Chicago were compared. All screen displays were presented, and data were collected, via a HyperCard stack designed for the purpose.

Sixty-two subjects determined that Geneva was the preferred font, with Boston a reasonably close second choice. Chicago was the last choice, with Courier not much more popular. Monaco and New York were somewhat below the midpoint of the top and bottom groups.

There is some indication that there may be an age effect with respect to one font (Monaco), but because of the design of the study (forced choice comparisons), the kind of analysis of the data required to make the determination was not possible. In any case, since Geneva held a firm first place, and Chicago and Courier were at the low end of the scale for both Younger and Older groups, the generalization that an instructional designer should carry away is simply to use Geneva whenever possible (with Boston as a second choice if a contrasting font is required) and avoid using Chicago and Courier.

REFERENCES

Dreyfus, J. (1985). A turning point in type design. *Visible Language, 19(1),* 11-22.

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),* 17-39.

Grabinger, R. S. (1985, January). *Relationships among text format variables in computer-generated text.* Paper presented at the Annual Conference of the Association for Educational Communications and Technology, Research and Theory Division, Las Vegas, NV.

Hartley, J. (1987). Designing electronic text: The role of print-based research. *Educational Communications and Technology Journal, 35(1),* 3-17.

Hooper, S., & Hannafin, M. J. (1986). Variables affecting the legibility of computer generated text. *Journal of Instructional Development,* 9(4), 22-28.

Maddox, M. E., Burnette, J. T, & Gutmann, J. C. (1977). Font comparisons for 5 x 7 dot matrix characters. *Human Factors, 19(1),* 89-93.

Poole, L. (1988). Quick tips: Answers to your questions. *MacWorld,* 5(5), 203.

Riley, T. M., & Barbato, G. J. (1978). Dot-matrix alphanumerics viewed under discrete element degradation. *Human Factors,* 20(4), 473—479.

Ross, R. T. (1934). Optimum orders for the presentation of pairs in the method of paired comparisons. *The Journal of Educational Psychology, 25,* 375-382.

Snyder, H. L., & Taylor, G. B. (1979). The sensitivity of response measures of alphanumeric legibility to variation in dot matrix display parameters. *Human Factors, 21(4),* 457-471.

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

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

AUTHOR

Earl R. Misanchuk is a Professor of Extension, The University of Sakatchewan, Saskatoon, Saskatchewan S7N OWO.